Estimation des Paramètres d'un Processus ARMA par Simulation MCMC à Sauts Réversibles.

S. Suparman* Michel Doisy†

Juin 2002

Résumé

Nous proposons dans cet article une méthode bayésienne pour l'estimation des paramètres des processus ARMA. Cette méthode s'appuie sur une simulation de type Monte Carlo par Chaine de Markov à Sauts Réversibles ([12]) pour une estimation conjointe de l'ensemble des paramètres à savoir : ordres p et q du processus ARMA et paramètres du modèle. Quatre jeux de données sont analysés et les résultats sont comparés à ceux obtenus par Brockwell et Davis grâce à leur logiciel PEST. Cette comparaison atteste la validité de notre approche qui travaille en aveugle, puisque les hyperparamètres introduits artificiellement dans toute méthode bayésienne sont actualisés au cours de l'Algorithme.

Mots clefs : Processus ARMA, Simulation de Monte Carlo par Chaines de Markov à Sauts Réversibles, Recuit Simulé.

Abstract

A new bayesian method for the estimation of the parameters of an ARMA Process is proposed in this article. Our approach is based on a Reversible Jumps Markov Chain Monte Carlo simulation ([12]) and provide both estimations of the orders p et q of the ARMA Process and of the parameters of the model. Four sets of data are analyzed and results are compared with those obtained by Brockwell et Davis through their programme PEST. More over, our technique is working *in blind,* hyperparameters of the a priori distributions being updated during the simulation.

Keywords: ARMA Processes, Reversible Jump Markov Chain Monte Carlo Simulation, Simulated Annealing.

*LSP-UPS, 118 route de Narbonne 31068 Toulouse Cedex 4, FRANCE (suparman@cict.fr)

†ENSEEIHT/TESA 2 Rue Camichel 31071 Toulouse Cedex 7, FRANCE (doisy@len7.enseeiht.fr)

1 Introduction

Les Processus Auto Régressif à Moyenne Mobile (ARMA) ont une grande importance pratique dans la modélisation des séries chronologiques stationnaires et une bibliographie importante existe pour l'étude de leur propriétés et l'estimation de leurs paramètres. Les deux références de base sur ce sujet sont Brockwell et Davis ([7]) et Box et Jenkins ([6]).

Un processus stochastique $(X_t, t \in \mathbf{Z})$ à valeurs réelles est un processus ARMA centré d'ordre p et q, noté ARMA(p,q) s' il est stationnaire et s'il vérifie une relation du type

$$X_t + \sum_{i=1}^{p} \phi_i^{(p)} X_{t-i} = \sum_{j=1}^{q} \theta_j^{(q)} Z_{t-j} + Z_t, \qquad t \in \mathbb{Z},$$

où $(p,q) \in \mathbb{N}^2$, $\boldsymbol{\phi}^{(p)} = \left(\phi_1^{(p)}, ..., \phi_p^{(p)}\right)' \in \mathbb{R}^p$, $\boldsymbol{\theta}^{(q)} = \left(\theta_1^{(q)}, ..., \theta_q^{(q)}\right)' \in \mathbb{R}^q$ (le symbole $'$ désigne la transposition du vecteur) et $Z_t, t \in \mathbb{Z}$ est un bruit blanc de moyenne 0 et de la variance $\sigma^2 \in \mathbb{R}^+$.

Il est donc caractérisé par deux familles de paramètres de nature distincte: les ordres p et q et les paramètres du modèle $\boldsymbol{\phi}^{(p)}$ et $\boldsymbol{\theta}^{(q)}$.

Pour p et q **fixés**, de nombreuses méthodes existent pour l'estimation des paramètres du modèle. Le chapitre 8 de Brockwell et Davis fait un tour d'horizon complet des estimateurs utilisés. Citons l'estimateur de Yule-Walker dans le cas $q = 0$, l'utilisation de l'Algorithme d'Innovation pour obtenir des estimations préliminaires dans le cas $q > 0$, et les estimateurs plus complets que ce sont les estimateurs des Moindres Carrés et du Maximum de Vraisemblance. Les propriétés asymptotiques et la détermination d'intervalles de confiance sont également connues permettant une connaissance très complète de ceux-ci.

Par contre, l'identification des valeurs appropriées pour p et q (*identification du modèle*) posent d'autres problèmes.

Une première approche consiste à utiliser le comportement de la Fonction d'Autocorrélation Partielle et de la Fonction d'Autocorrélation. En effet, si la partie AR est d'ordre p alors la Fonction d'Autocorrélation Partielle s'annule pour $k > p$ (resp si la partie MA est d'ordre q c'est la Fonction d'Autocorrélation qui s'annule pour $l > q$). La difficulté dans l'utilisation de cette méthode est que ces deux fonctions sont *estimées* à partir des données et qu'il est délicat de déterminer le seuil du *décrochage* (*cut off*). L'étude des propriétés asymptotiques des estimateurs de ces fonctions palie en partie à cette difficulté en fournissant un intervalle de confiance pour obtenir un tel seuil ([7] p. 243 et p.247). Neanmoins, comme le soulignent les auteurs, si cette approche convient pour déterminer l'ordre d'un processus "purement" Auto Régresssif ou d'un processus "purement" MA, elle devient beaucoup plus délicate dans le cas d'un processus mixte (*cf.* l'exemple des données du lac de Huron que nous traiterons dans la partie simulations).

Une autre approche importante consiste à utiliser des *critères pénalisés* du type Akaike (AICC, AIC, BIC). Ces critères tiennent compte à la fois de l'adéquation aux données mais en même temps évitent une sur-paramètrisation du modèle en pénalisant l'augmentation des ordres (*cf* : Akaike (1973), Hurvich *et al.* (1989), et Schwarz (1978)). Ils utilisent dans leur calcul les Estimateurs du Maximum de Vraisemblance des paramètres du modèle. Le principe est alors de déterminer le modèle qui minimise le critère et donc nécessite son calcul pour l'ensemble des modèles candidats (même si des techniques de réduction existent!).

Enfin une analyse basée sur l'existence de relations de récurrence de l'autocorrélation peut être trouvée dans Beguin *et al.* (1980) par exemple, mais elle reste délicate à mettre en oeuvre.

Nous adoptons dans ce travail une approche *bayésienne*: ordres et paramètres du modèle sont vues comme des variables aléatoires et pourvues de lois à priori. Dans un tel cadre les estimations sont basées sur la *loi à postériori* associée au modèle. La complexité de celle-ci rend impossible l'obtention d'une expression explicite de l'Estimateur du Maximum de Vraisemblance des paramètres, par exemple. Nous avons alors recours à une simulation de type Monte Carlo par Chaine de Markov. Rappelons brièvement qu'il s'agit de construire une chaine de Markov, à valeurs dans l'espace des paramètres et qui admette comme distribution stationnaire, la loi à postériori dont nous souhaitons étudier les propriétés. Comme les ordres p et q font partie des paramètres à estimer, l'espace d'états de la chaine se présente, en ce qui concerne les paramètres $\boldsymbol{\phi}^{(p)}$ et $\boldsymbol{\theta}^{(q)}$, comme une réunion de sous espaces de dimension différente *i.e.* comme un sous-ensemble de

$$\bigcup_{\substack{p=0,\dots,p_{\max}\\ q=0,\dots,q_{\max}}} \mathbb{R}^p \times \mathbb{R}^q$$

Pour utiliser un algorithme de type Metropolis-Hastings avec un tel espace d'états il est nécessaire de définir un noyau de transition qui permette à la chaine d'effectuer des sauts entre espace de *dimension différente,* tout en restant *irréductible.* Nous utilisons la solution, maintenant bien connue, proposée par P. Green ([12]) dite MCMC à sauts réversibles.

La dernière difficulté provient des conditions de *causalité* et d'*inversibilité* (*cf.* paragraphe 2) des processus ARMA utilisés: le vecteur $\boldsymbol{\phi}^{(p)}$ (resp. $\boldsymbol{\theta}^{(q)}$) doit prendre ses valeurs dans un sous ensemble $\mathcal{C}_p$ (resp. $\mathcal{I}_q$) de $\mathbb{R}^p$ (resp $\mathbb{R}^q$) pour être causal (resp inversible). En général les ensembles $\mathcal{C}_p$ et $\mathcal{I}_q$ n'ont pas de caractérisation simple, il est alors impossible de simuler un tirage de points dans ces ensembles. Certains auteurs (Liu (1995), et Shaarawy *et al.* (1985)) ont choisi de ne pas tenir compte de ces conditions, mais ils travaillent avec p et q fixés.

Suivant une idée introduite par Barbieri et O'Hagan ([2]) dans un travail non publié, pour tenir compte des conditions de causalité et d'inversibilité dans les simulations, nous avons adopté la reparamètrisation introduite par Monahan ([19]). Elle s'appuie sur les notions d'*Autocorrélation Partielle* et d'*Autocorrélation Partielle Inverse.* Cette nouvelle paramètrisation met le domaine $\mathcal{C}_p \times \mathcal{I}_q$ en bijection avec $]-1,+1[^p \times]-1,+1[^q$ dans lequel il est facile de tirer des points de façon aléatoire. La contre partie est qu'il n'existe pas d'expression analytique simple liant ces deux familles de paramètres, les calculs s'effectuant de proche en proche (*cf.* paragraphe 2.4). Dans ce nouvel espace de paramètres nous définissons les sauts permettant de faire tourner un algorithme MCMC à sauts réversibles. Enfin pour éviter que les hyperparamètres, introduits de façon artificielle dans les lois à priori n'influent trop sur le comportement de l'algorithme, nous mettons des lois hyper à priori sur ceux-ci et nous les actualisons au cours de l'algorithme.

L'algorithme est utilisé pour obtenir l'Estimateur du Maximum de Vraisemblance Marginal à Postériori pour p et q, puis les Moyennes Marginales à Postériori pour les paramètres du modèle. Afin d'utiliser plus complétement les sorties de l'algorithme nous donnons également l'estimation par noyau des Densités Marginales à Postériori de $\boldsymbol{\phi}^{(p)}$ et $\boldsymbol{\theta}^{(q)}$. Enfin l'adjonction d'un schéma de descente de température au noyau de la chaine, permet de mettre facilement en place un Algorithme de Recuit Simulé qui fournit une estimation du Maximum de Vraisemblance pour l'ensemble des paramètres.

Nous avons choisi de tester cette méthode sur quatre exemples et de comparer nos résultats avec ceux proposés par Brockwell et Davis au travers de leur logiciel PEST. Deux sont des exemples de données réelles, les deux autres des données simulées.

Dans les parties 2 et 3 de cet article nous rappelons les notions que nous utiliserons concernant les processus ARMA ainsi que les méthodes "classiques" d'estimation des paramètres. Dans la partie 4 nous décrivons la méthode de simulation MCMC à sauts réversibles. Nous détaillons dans la partie suivante notre algorithme et la nature des estimations effectuées. Enfin, la dernière partie regroupe nos quatre exemples de simulations, les résultats et commentaires.

2 Rappels sur les processus ARMA

Dans ce paragraphe, nous rappelons les notions essentielles concernant les processus ARMA(p,q). Nous renvoyons à [6] ou [7] pour l'étude détaillé de ceux-ci.

2.1 Notations

Rappelons tout d'abord qu'un processus stochastique $(X_t, t \in \mathbf{Z})$ à valeurs réelles est *stationnaire du second ordre* s'il est carré intégrable, $E(X_t)$ ne dépend pas de t et

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

ne dépend que de h $(\forall t, h \in \mathbb{Z})$. Dans la suite, on employera le terme "stationnaire" pour "stationnaire du secondre ordre".

Nous adoptons la définition des processus ARMA de Brockwell et Davis [7]. Un processus stochastique $(X_t, t \in \mathbf{Z})$ à valeurs réelles est un processus ARMA centré d'ordre p et q, noté ARMA(p,q) s' il est stationnaire et il vérifie une relation du type

$$X_t + \sum_{i=1}^{p} \phi_i^{(p)} X_{t-i} = \sum_{j=1}^{q} \theta_j^{(q)} Z_{t-j} + Z_t, \qquad t \in \mathbb{Z}, \tag{1}$$

où $(p,q) \in \mathbb{N}^2$, $\boldsymbol{\phi}^{(p)} = \left(\phi_1^{(p)}, ..., \phi_p^{(p)}\right)' \in \mathbb{R}^p$, $\boldsymbol{\theta}^{(q)} = \left(\theta_1^{(q)}, ..., \theta_q^{(q)}\right)' \in \mathbb{R}^q$ (le symbole $'$ désigne la transposition du vecteur) et $(Z_t, t \in \mathbb{Z})$ est un bruit blanc de moyenne 0 et de la variance $\sigma^2 \in \mathbb{R}^+$. On supposera dans la suite que le bruit blanc est un processus gaussien[1]. Si $\boldsymbol{\phi}(.)$ et $\boldsymbol{\theta}(.)$ désignent les deux polynômes

$$\boldsymbol{\phi}(z) = 1 + \phi_1^{(p)} z + \phi_2^{(p)} z^2 + ... + \phi_p^{(p)} z^p$$

et

$$\boldsymbol{\theta}(z) = 1 + \theta_1^{(q)} z + \theta_2^{(q)} z^2 + ... + \theta_q^{(q)} z^q$$

et B l'*opérateur retard* $(B^j X_t = X_{t-j}\ , t \in \mathbb{Z})$, on peut alors réecrire l'équation (1) sous la forme

$$\boldsymbol{\phi}(B) X_t = \boldsymbol{\theta}(B) Z_t. \tag{2}$$

On sait par ailleurs qu'un processus défini par l'équation (1) est stationnaire si et seulement si les racines complexes du polynôme $\boldsymbol{\phi}(z)$ ne sont pas de module 1.

[1] Les calculs dans le cadre gaussien jouant un rôle essentiel, même lorsque le bruit blanc n'est pas gaussien, il est usuel de faire cette hypothèse.

Le processus $(X_t, t \in \mathbf{Z})$, est *causal* si X_t peut s'exprimer en fonction des Z_s avec $s \leq t$. On a aussi la caractérisation classique suivante de la causalité: $(X_t, t \in \mathbf{Z})$ est causal si et seulement les racines complexes du polynôme $\boldsymbol{\phi}(z)$ sont de module strictement supérieur à 1.

La notion d'*inversibilité* est la notion duale de la précédente. Le processus $(X_t, t \in \mathbf{Z})$ est *inversible* si Z_t peut s'exprimer à partir des X_s pour $s \leq t$. On a une caractérisation analogue portant sur les racines du polynôme $\boldsymbol{\theta}(z)$: $(X_t, t \in \mathbf{Z})$ est inversible si et seulement les racines complexes du polynôme $\boldsymbol{\theta}(z)$ sont de module supérieur à 1.

Définition 1 *Pour p et q deux entiers non nuls donnés, on appellera domaine de causalité (resp. domaine d'inversibilité) l'ensemble*

$$\mathcal{C}_p = \left\{ \boldsymbol{\phi}^{(p)} = \left(\phi_1^{(p)}, ..., \phi_p^{(p)}\right)' \in \mathbb{R}^p \quad t.q. \quad \boldsymbol{\phi}(z) \neq 0 \quad \forall z \in \mathbb{C}\,, |z| \leq 1 \right\}$$

resp.

$$\mathcal{I}_q = \left\{ \boldsymbol{\theta}^{(q)} = \left(\theta_1^{(q)}, ..., \theta_q^{(q)}\right)' \in \mathbb{R}^q \quad t.q. \quad \boldsymbol{\theta}(z) \neq 0 \quad \forall z \in \mathbb{C}\,, |z| \leq 1 \right\}$$

Des conditions plus simples de causalite et inversibilité seront données dans §2.4. Dans ce qui suit, on supposera que les processus ARMA(p,q) étudiés vérifient les conditions de causalité et invertibilité.

2.2 Fonction de vraisemblance exacte

Soient $\mathbf{x}_n' = (x_1, ..., x_n)$ le vecteur des observations d'un processus ARMA(p,q) et $\mathbf{X}_n' = (X_1, ..., X_n)$ le vecteur aléatoire correspondant. On note Γ_n la matrice de covariance associée soit $\Gamma_n = E(\mathbf{X}_n \mathbf{X}_n')$. Comme le bruit blanc est supposé gaussien[2], la *vraisemblance exacte*[3] est donnée par

$$\ell_e(\mathbf{X}_n \mid (p,q), \phi^{(p)}, \theta^{(q)}, \sigma^2) = (2\pi)^{-\frac{n}{2}} \det(\Gamma_n)^{-\frac{1}{2}} \exp -\frac{1}{2} \mathbf{X}_n' \Gamma_n^{-1} \mathbf{X}_n. \tag{3}$$

Une autre forme de cette vraisemblance est plus commode dans les problèmes d'estimation (cf.[8] p. 158) . Notons $\widehat{X}_j$ $(j = 1, ..., n)$ la projection orthogonale de X_j sur $X_1, ..., X_{j-1}$ (avec par convention $\widehat{X}_1 = 0$). Matriciellement on peut écrire

$$\mathbf{X}_n - \widehat{\mathbf{X}}_n = A_n \mathbf{X}_n$$

la matrice A_n étant triangulaire inférieure avec des facteurs 1 sur la diagonale. On en déduit

$$\mathbf{X}_n = B_n \left(\mathbf{X}_n - \widehat{\mathbf{X}}_n \right)$$

si $B_n = A_n^{-1}$ ou encore, en posant $C_n = B_n - I_n$ $\widehat{\mathbf{X}}_n = C_n \left(\mathbf{X}_n - \widehat{\mathbf{X}}_n \right)$. La matrice C_n est triangulaire inférieure, avec des facteurs 0 sur la diagonale. La formule précédente donne l'ensemble des relations récursives suivantes

$$\widehat{X}_{k+1} = \begin{cases} 0 & \text{si } k = 0, \\ \sum_{j=1}^{k} c_{t,j} (X_{k+1-j} - \widehat{X}_{k+1-j}) & \text{si } k = 1, 2, .., n-1. \end{cases} \tag{4}$$

[2] Brockwell et Davis soulignent (p. 255) que même lorsque le processus n'est pas gaussien, cette *vraisemblance gaussienne* sert comme mesure d'adéquation de la matrice de covariance aux données.

[3] Dans ce qui suit on écrit les quantités d'intérêt en termes de variables aléatoires *i.e.* en fonction des X_n. Dans les estimations, on reviendra aux observations *i.e.* aux $x_1, \ldots, x_n$.

De plus, par définition des projections orthogonales, il est clair que les variables aléatoires $X_j - \widehat{X_j}$ ne sont pas corrélées. En notant

$$v_j = \frac{1}{\sigma^2} E(X_{j+1} - \widehat{X}_{j+1})^2 \, (j = 0, \cdots, n-1)$$

l'expression de la vraisemblance devient

$$\begin{aligned} \ell e(\mathbf{X}_n \mid (p,q), \phi^{(p)}, \theta^{(q)}, \sigma^2) &= (2\pi\sigma^2)^{-n/2} (\upsilon_0 ... \upsilon_{n-1})^{-1/2} \\ &\quad \exp\left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^{n} (X_j - \widehat{X}_j)^2 / \upsilon_{j-1} \right\}. \end{aligned} \tag{5}$$

Dans cette formule, les $\widehat{X}_j$ ou ce, qui est équivalent par (4) les c_{ij} d'une part et les v_j d'autre part, peuvent être obtenus par l'*algorithme d'innovation* ([8] p. 71). Soit $V_n = (V_n(i,j))_{i,j} = \frac{1}{\sigma^2}\Gamma_n$.

Algorithme 2 (Algorithme d'Innovation) *Les coefficients $c_{n,1}, ..., c_{n,n}$ dans (4) peuvent être calculés récursivement par les équations*

$$\begin{aligned} \upsilon_0 &= V_n(1,1), \\ c_{n,n-k} &= \upsilon_k^{-1} (V_n(n+1, k+1) - \sum_{j=0}^{k-1} c_{k,k-j} c_{n,n-j} \upsilon_j), \qquad 0 \leq k < n, \end{aligned}$$

et

$$\upsilon_n = V_n(n+1, n+1) - \sum_{j=0}^{k-1} c_{n,n-j}^2 \upsilon_j.$$

Il est important de remarquer, que cet algorithme ne fait par intervenir σ^2.

2.3 Fonction de vraisemblance approchée

L'expression fonctionnelle de vraisemblance exacte (3) apparemment simple, fait en réalité intervenir de façon complexe les paramètres $\phi^{(p)}$et $\theta^{(q)}$ et est donc difficilement utilisable. Plusieurs approximations de cette fonction de vraisemblance ont été proposées dans la littérature ([6],[18], [25]). Nous utiliserons dans la suite celle développée par S.I. Lui ([18]) et S. Shaarawy et L. Broemeling ([25]).

Dans cette approximation de la vraisemblance d'un modèle ARMA(p,q), on considère le vecteur $\mathbf{x}'_p = (x_1, ..., x_p)$ comme vecteur des *valeurs initiales* et on prend $z_{p+1-q} = ... = z_p = 0$. Alors, la fonction de vraisemblance approchée s'écrit

$$\begin{aligned} \ell_a(\mathbf{X}_n \mid (p,q), \phi^{(p)}, \theta^{(q)}, \sigma^2) &\simeq \left(2\pi\sigma^2\right)^{-\frac{n-p}{2}} \\ &\quad \exp -\tfrac{1}{2\sigma^2} \sum_{t=p+1}^{n} e^2(t, p, q, \phi^{(p)}, \theta^{(q)}) \end{aligned} \tag{6}$$

où pour tout $t = 1, ..., n$

$$e(t, p, q, \phi^{(p)}, \theta^{(q)}) = X_t + \sum_{i=1}^{p} \phi_i X_{t-i} - \sum_{j=1}^{q} \theta_i \widehat{Z}_{t-i}. \tag{7}$$

Dans (7) les valeurs $\widehat{Z}_t$ sont des *valeurs estimées* du bruit, l'estimation est obtenu par l'algorithme ([18], [25])

$$\widehat{Z}_t = X_t + \sum_{i=1}^{p} \phi_i X_{t-i} - \sum_{j=1}^{q} \theta_i \widehat{Z}_{t-i}, \qquad t = p+1, ..., n.$$

2.4 Reparamètrisation

Dans une méthode d'estimation de paramètres de type Monte Carlo, on explore aléatoirement l'espace des paramètres. Ici, l'estimation des paramètres $\phi^{(p)}$et $\theta^{(q)}$ d'un processus ARMA causal et inversible requiert de tirer des points dans les domaines de causalité $\mathcal{C}_p$ et $\mathcal{T}_p$. Exprimés sous leur forme initiale en termes de $\phi^{(p)}$ et $\theta^{(q)}$, ces domaines n'ont pas une caractérisation simple. Pour résoudre ce problème, nous utilisons la reparamètrisation introduite par Monahan ([19]). Celle-ci s'appuie sur les notations d'*Autocorrélation Partielle* et d'*Autocorrélation Partielle Inverse* définies précisement dans ([3], [5]). Nous rappelons dans ce qui suit l'aspect qui nous interesse, à savoir le passage des anciens paramètres $\phi^{(p)}$ et $\theta^{(q)}$ à ces nouveaux paramètres.

Si p et q sont les ordres du processus, on notera $r^{(p)} = (r_1, ..., r_p)$ le vecteur des autocorrelations partielles et $\rho^{(q)} = (\rho_1, ..., \rho_q)$ le vecteur des autocorrelations partielles inverses. Ces vecteurs sont contenus dans les cubes $]-1, +1[^p$ et $]-1, +1[^q$.

Résultat

Pour les ARMA(p,q) causaux et inversbles les transformations qui associent à

$$F : \phi^{(p)} = (\phi_1^{(p)}, ..., \phi_p^{(p)}) \in \mathcal{C}_p \mapsto r^{(p)} = (r_1, ..., r_p) \in]-1, +1[^p \tag{8}$$

et

$$G : \theta^{(q)} = (\theta_1^{(q)}, ..., \theta_q^{(q)}) \in \mathcal{I}_q \mapsto \rho^{(q)} = (\rho_1, ..., \rho_q) \in]-1, +1[^q \tag{9}$$

sont bijectives.

Par contre le lien entre $\phi^{(p)}$ et $r^{(p)}$ d'une part et $\theta^{(q)}$ et $\rho^{(q)}$ d'autre part, ne posséde pas d'expression analytique simple. Par exemple pour passer de $r^{(p)}$ à $\phi^{(p)}$ on utilise un tableau triangulaire du type suivant

$$\begin{array}{lllll}
\phi_1^{(1)} = r_1 & & & & \\
\phi_1^{(2)} & \phi_2^{(2)} = r_2 & & & \\
\phi_1^{(3)} & \phi_2^{(3)} & \phi_3^{(3)} = r_3 & & \\
\vdots & \vdots & \vdots & \ddots & \\
\phi_1^{(p)} & \phi_2^{(p)} & \phi_3^{(p)} & \cdots & \phi_p^{(p)} = r_p
\end{array}$$

avec, au sein de ce tableau, les relations

$$\left\{ \begin{array}{lcllcl}
\phi_1^{(1)} & = & r_1 & & & \\
\phi_i^{(k)} & = & \phi_i^{(k-1)} + r_k \phi_{k-i}^{(k-1)}, & i & = & 1, ..., k-1 \\
\phi_k^{(k)} & = & r_k, & k & = & 2, ..., p
\end{array} \right. \tag{10}$$

Ainsi pour obtenir le vecteur $\phi^{(p)}$ à partir de la donnée du vecteur $r^{(p)}$, on doit calculer, de proche en proche les termes de toutes les lignes grâce aux relations (10). Inversement pour obtenir $r^{(p)}$ à partir de $\phi^{(p)}$, on calcule tous les termes du tableau triangulaire, à partir de la dernière ligne du tableau et aux relations inverses de (10)

$$\phi_i^{(k-1)} = (\phi_i^{(k)} - r_k \phi_{k-i}^{(k)})/(1 - r_k^2), \qquad i = 1, ..., k-1 \text{ et } k = 2, ..., p,$$

ensuite le valeur $r^{(p)}$ est donné par $r_k = \phi_k^{(k)}$ $(k = 1, ..., p)$. Le lien entre $\theta^{(q)}$ et $\rho^{(q)}$ est du même type.

Si on n'a pas d'expresssion analytique directe permettant de passer des anciens paramètres aux nouveaux, le jacobien de ces transformations est connu (*cf.* [15])

Résultat

Le Jacobien des applications F^{-1} et G^{-1} est donné par les formules

$$J_{F^{-1}}(r^{(p)}) = \prod_{k=2}^{p} (1 + r_k)^{[k/2]} (1 - r_k)^{[(k-1)/2]}$$

$$(\text{resp.} J_{G^{-1}}(q^{(q)}) = \prod_{h=2}^{q} (1 + \rho_h)^{[h/2]} (1 - \rho_h)^{[(h-1)/2]} \ )$$

pour tout $p \geq 2$ (resp. $q \geq 2$) et est égal à 1 pour $p = 1$ (resp. $q = 1$), ici $[x]$ la partie entière de x. Grace à cette reparamètrisation, les caractérisations $\phi^{(p)} \in \mathcal{C}_p$ et $\theta^{(q)} \in \mathcal{I}_q$ sont équivalentes à $|r_k| < 1$ $(k = 1, ..., p)$ et $|\rho_h| < 1$ $(h = 1, ..., q)$.

On peut alors écrire la fonction de vraisemblance (6) en fonction des nouveaux paramètres

$$\begin{aligned} \ell_a(\mathbf{X}_n \mid (p,q), r^{(p)}, \rho^{(q)}, \sigma^2) &= \left(2\pi\sigma^2\right)^{-\frac{n-p}{2}} \\ &\quad \exp -\tfrac{1}{2\sigma^2} \sum_{t=p+1}^{n} e^2(t, p, q, F^{-1}(r^{(p)}), G^{-1}(\rho^{(q)})) \end{aligned} \tag{11}$$

Pour des valeurs p et q petites, il est possible de donner l'expression analytique de (11) comme l'exemple ci-dessous ([3]).

Exemple 3 *Dans le cas où $p = 2$ et $q = 0$, on obtient*

$$\begin{aligned} \ell_a(\mathbf{X}_n \mid r^{(2)}, \sigma^2) &= \tfrac{1}{(2\pi\sigma^2)^{n/2}} (1 - r_1^2)^{1/2} (1 - r_2^2) \\ &\quad \exp -\tfrac{1}{2\sigma^2} g(r^{(2)}) \times (1 - r_2)(r_2 + 1)^{[1/2]} \end{aligned}$$

où la fonction g définie par

$$\begin{aligned} g(r^{(2)}) &= \sum_{i=1}^{n} x_i^2 + r_1^2 (1 - r_2)^2 \sum_{i=2}^{n-1} x_i^2 + r_2^2 \sum_{i=3}^{n-2} x_i^2 \\ &\quad -2r_1(1 - r_2) \sum_{i=2}^{n} x_{i-1} x_i - 2r_2 \sum_{i=3}^{n} x_{i-2} x_i \\ &\quad +2r_1 r_2 (1 - r_2) \sum_{i=3}^{n-1} x_{i-1} x_i. \end{aligned}$$

Par contre dans le cas général, on devra calculer de proche en proche, les termes du tableau triangulaire, comme il a été indiqué plus haut.

Pour la méthode bayésienne que nous présentons dans §4, on travaillera avec les nouveaux paramètres $r^{(p)}$ et $\rho^{(q)}$.

3 Méthodes Classiques d'estimation des ordres

Les méthodes dites *classiques* - par opposition aux méthodes bayésiennes que nous allons développer - d'estimation des ordres p et q d'un processus ARMA, utilisent en premier lieu l'estimation des autres paramètres du modèle, $\phi^{(p)}$, $\theta^{(q)}$, et σ^2. Nous commençons donc par rappeler les différentes méthodes d'estimation de ceux-ci.

3.1 Estimation des paramètres

Lorsque p et q sont fixés, il existe différentes méthodes pour estimer des paramètres $\phi^{(p)}$, $\theta^{(q)}$, et σ^2 du modèle ARMA(p,q). La technique la plus usuelle est d'estimer ces paramètres par l'estimateur du maximum de vraisemblance, c'est-à-dire des valeurs de $(\phi^{(p)}, \theta^{(q)}, \sigma^2) \in \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^+$ qui maximisent la fonction de vraisemblance (3). Pour cela, on annule ses dérivées partielles par rapport aux paramètres $\phi^{(p)}$, $\theta^{(q)}$, et σ^2. En rappellant que $\widehat{X}_j$ et υ_j sont indépendants de σ^2, on en déduit que les Estimateurs du Maximum de Vraisemblance $\widehat{\phi}^{(p)}$, $\widehat{\theta}^{(q)}$, et $\widehat{\sigma}^2$ vérifient

$$\widehat{\sigma}^2 = n^{-1} S_x(\widehat{\phi}^{(p)}, \widehat{\theta}^{(q)}) \tag{12}$$

où

$$S_x(\widehat{\phi}^{(p)}, \widehat{\theta}^{(q)}) = \sum_{j=1}^{n} (X_j - \widehat{X}_j)^2 / \upsilon_{j-1} \tag{13}$$

et $\widehat{\boldsymbol{\phi}}^{(p)}$, $\widehat{\theta}^{(q)}$ sont les valeurs de $\boldsymbol{\phi}^{(p)}$, $\theta^{(q)}$ qui minimisent

$$l_r(\boldsymbol{\phi}^{(p)}, \theta^{(q)}) = \ln(n^{-1} S_x(\boldsymbol{\phi}^{(p)}, \boldsymbol{\theta}^{(q)})) + n^{-1} \sum_{j=1}^{n} \ln(\upsilon_{j-1}). \tag{14}$$

La difficulté est que l'on n'a pas de forme explicite pour la minimisation de (14). On a donc recours à des méthodes numériques d'optimisation non linéaire avec comme problème crucial le choix correcte de la valeur initiale, qui conditionne la rapidité de l'algorithme. Plusieurs méthodes sont proposées par Brockwell et Davis, selon le type de processus: l'Algorithme de Hannan-Rissanen ou l'Algorithme d'Innovation.

3.2 Identification des ordres

En ce qui concerne l'estimation des ordres p et q d'un processus ARMA(p,q), on peut distinguer d'une part les méthodes basées sur l'existence de la relation de récurrence de la covariance du modèle ([4], [22]) et d'autre part, les critères comme AICC, AIC, BIC (cf. [1], [14], [26]).

Nous rappelons l'idée de l'un de ces critères, la version biais-corrigé du critère AIC, noté , AICC proposée dans ([14]). Les autres critères sont similaires à celui d'AICC avec différentes pénalisations. Rappelons que si $\mathbf{X}_n$ est un vecteur aléatoire de dimension n dont la loi de probabilités appartient à une famille de lois, paramètrée par un ensemble Ψ, soit $\mathcal{F} = \{f(. \mid \psi), \psi \in \Psi)\}$ la *divergence Kullback-Leibler* entre $f(. \mid \psi)$ et $f(. \mid \varphi)$, notée $d(\psi \mid \varphi)$ est donnée par

$$d(\psi \mid \varphi) = \Delta(\psi \mid \varphi) - \Delta(\varphi \mid \varphi)$$

où $\Delta(\psi \mid \varphi)$ est l'*indice Kullback-Leibler* de $f(. \mid \psi)$ relative à $f(. \mid \varphi)$

$$\Delta(\psi \mid \varphi) = \int_{\mathbb{R}^n} -2 \ln(f(\mathbf{x}_n \mid \psi)) f(\mathbf{x}_n \mid \varphi) d\mathbf{x}_n$$

L'utilisation de cette distance pour estimer les ordres d'un processus ARMA gaussien se faite de la façon suivante. Supposons que l'ordre soit égal à (p,q) et soit $(\widehat{\phi}^{(p)}, \widehat{\theta}^{(q)}, \widehat{\sigma}^2)$ l'Estimateur de Maximum de

Vraisemblance de $(\phi^{(p)}, \theta^{(q)}, \sigma^2)$ fondé sur $\mathbf{x}_n$ obtenu par (14) et (12). Soit $y_1, ..., y_n$ une autre réalisation du modèle vrai avec le vecteur de paramètre $(\phi^{(p)}, \theta^{(q)}, \sigma^2)$, et *indépendante* de $x_1, ..., x_n$. Notons

$$le_y(\widehat{\phi}^{(p)}, \widehat{\theta}^{(q)}, \widehat{\sigma}^2) = (2\pi\widehat{\sigma}^2)^{-n/2}(\upsilon_0 ... \upsilon_{n-1})^{-1/2} \exp\left\{-\frac{1}{2\widehat{\sigma}^2} S_y(\widehat{\phi}^{(p)}, \widehat{\theta}^{(q)})\right\},$$

avec

$$S_y(\widehat{\phi}^{(p)}, \widehat{\theta}^{(q)}) = \sum_{j=1}^{n} (y_j - \widehat{y}_j)^2 / \upsilon_{j-1},$$

puis

$$le_x(\widehat{\phi}^{(p)}, \widehat{\theta}^{(q)}, \widehat{\sigma}^2) = (2\pi\widehat{\sigma}^2)^{-n/2}(\upsilon_0 ... \upsilon_{n-1})^{-1/2} \exp\left\{-\frac{1}{2\widehat{\sigma}^2} S_x(\widehat{\phi}^{(p)}, \widehat{\theta}^{(q)})\right\}.$$

avec

$$S_x(\widehat{\phi}^{(p)}, \widehat{\theta}^{(q)}) = \sum_{j=1}^{n} (x_j - \widehat{x}_j)^2 / \upsilon_{j-1}.$$

Puisque

$$\widehat{\sigma}^2 = n^{-1} S_x(\widehat{\phi}^{(p)}, \widehat{\theta}^{(q)}),$$

on a

$$-2 \ln le_y(\widehat{\phi}^{(p)}, \widehat{\theta}^{(q)}, \widehat{\sigma}^2) = -2 \ln le_x(\widehat{\phi}^{(p)}, \widehat{\theta}^{(q)}, \widehat{\sigma}^2) + \widehat{\sigma}^{-2} S_y(\widehat{\phi}^{(p)}, \widehat{\theta}^{(q)}) - n.$$

Ceci entraîne que

$$\begin{aligned} \Delta(\widehat{\phi}^{(p)}, \widehat{\theta}^{(q)}, \widehat{\sigma}^2 \mid \phi^{(p)}, \theta^{(q)}, \sigma^2) &= E(-2 \ln le_y(\widehat{\phi}^{(p)}, \widehat{\theta}^{(q)}, \widehat{\sigma}^2)) \\ &= -2 \ln le_x(\widehat{\phi}^{(p)}, \widehat{\theta}^{(q)}, \widehat{\sigma}^2) + E(\frac{S_y(\widehat{\phi}^{(p)}, \widehat{\theta}^{(q)})}{\widehat{\sigma}^2}) - n. \end{aligned} \tag{15}$$

Sous l' hypothèse que tous les modèles candidats sont gaussiens, Brockwell *et al.* ([7], page 303-304) ont montré que

$$E(\frac{S_y(\widehat{\phi}^{(p)}, \widehat{\theta}^{(q)})}{\widehat{\sigma}^2}) \simeq \frac{2(p+q+1)n}{n-p-q-2} + n.$$

Ainsi la quantité

$$-2 \ln le_x(\widehat{\phi}^{(p)}, \widehat{\theta}^{(q)}, \widehat{\sigma}^2) + \frac{2(p+q+1)n}{n-p-q-2}$$

est un estimateur sans biais approché de l'indice Kullback-Leibler $\Delta(\widehat{\phi}^{(p)}, \widehat{\theta}^{(q)}, \widehat{\sigma}^2 \mid \phi^{(p)}, \theta^{(q)}, \sigma^2)$ donné par (15).

On choisit alors les valeurs p et q du modèle ajusté qui minimise $AICC(\widehat{\phi^{(p)}}, \widehat{\theta^{(q)}})$ en notant

$$AICC(\phi^{(p)}, \theta^{(q)}) = -2 \ln le_x(\phi^{(p)}, \theta^{(q)}, \frac{S_x(\phi^{(p)}, \theta^{(q)})}{n}) + \frac{2(p+q+1)n}{n-p-q-2}. \tag{16}$$

Autrement dit, on choisit p et q conduisant à l'estimateur des divergences Kullback-Leibler $\Delta(\widehat{\phi}^{(p)}, \widehat{\theta}^{(q)}, \widehat{\sigma}^2 \mid \phi^{(p)}, \theta^{(q)}, \sigma^2)$ minimum. Pour le critère AIC, c'est ce dernier terme de pénalisation $\dfrac{2(p+q+1)n}{n-p-q-2}$ qui deviendra $2(p+q+1)$**.** Enfin pour le critère BIC, il est défini par

$$BIC(\sigma^2) = (n-p-q)\ln le_x(\frac{n\sigma^2}{n-p-q}) + n(1+\ln\sqrt{2\pi}) + (p+q)\ln\left[\frac{\sum_{i=1}^{n} x_i^2 - n\sigma^2}{p+q}\right].$$

4 Approche Bayésienne et Methodes MCMC

Dans la méthode classique précédente, les estimation des paramètres ($\phi^{(p)}$, $\theta^{(q)}$, σ^2) et des ordres (p,q) d'une processus ARMA(p,q) ne sont pas obtenus simultanément. Autrement dit, à partir d'une observation $\mathbf{x}_n$ nous estimons d'abord les ordres (p,q) et ensuite à (p,q) fixés nous estimons les paramètres ($\phi^{(p)}$, $\theta^{(q)}$, σ^2).

Nous proposons dans ce qui suit une méthode bayésienne pour estimer conjointement tous les paramètres (p,q) et ($\phi^{(p)}$, $\theta^{(q)}$, σ^2). Comme indiqué dans § 2.4., pour cet approche, nous travaillons avec les nouveaux paramètres ($\mathbf{r}^{(p)}$, $\boldsymbol{\rho}^{(q)}$). Nous désignons par $\mathcal{M}_{p,q}$ l'ensemble des processus ARMA(p,q) et considérons la famille des modèles

$$\mathcal{M} = \{\mathcal{M}_{p,q}; p = 0,1,...,p_{\max}, q = 0,1,...,q_{\max}\}.$$

A partir du vecteur d'observations $\mathbf{x}_n' = (x_{1,} \ldots x_n)$ nous souhaitons choisir un modèle dans la famille $\mathcal{M}$ puis estimer les paramètres du modèle, ($\mathbf{r}^{(p)}$, $\boldsymbol{\rho}^{(q)}$, σ^2). Nous avons donc un problème de *choix de modèle*.

L'estimation bayésienne de (p,q) et ($\mathbf{r}^{(p)}$, $\boldsymbol{\rho}^{(q)}$, σ^2) s'appuie sur sa loi *à posteriori* conjointe qui se décompose en produit de la fonction de vraisemblance et de la loi à priori. Dans ce type de problème, la loi à postériori a une forme complexe qui ne permet pas un traitement analytique directe et donc qui ne fournit pas d'expression explicite des estimateurs. Ceux-ci seront obtenus en simulant selon la loi à postériori. Là encore, la forme complexe de cette loi oblige à recourir à des techniques de simulation indirectes: nous utiliserons la méthode de Monte Carlo par Chaîne de Markov (MCMC) à sauts réversibles introduit par P. Green ([12]).

4.1 Approche Bayésienne

Dans cette approche bayésienne, le vecteur de paramètre inconnu $\mathcal{H}_1 = \left(p, q, \mathbf{r}^{(p)}, \boldsymbol{\rho}^{(q)}, \sigma^2\right)'$ est considéré comme une variable aléatoire, avec une distribution *à priori* connue qui traduit notre degré de confiance dans les différentes valeurs de ces quantités. Ces lois, dépendent souvent elles-mêmes de paramètres appelés *hyperparamètres*, rajoutés arbitrairement au modèle. Pour augmenter la robusteuse de la méthode, nous mettrons également des lois, dites hyper à priori sur ces paramètres ([23] pp. 293-302) . Nous proposons donc un modèle *bayésien hierarchique*. Nous précisons dans ce qui suit les lois à priori et hyper à priori choisies.

Dans notre travail, nous utilisons la distribution Binomiale de paramètre $0 < \lambda < 1$ (resp. $0 < \mu < 1$) pour la distribution à priori pour l'ordre p (resp. q)

$$\pi(p;\lambda) = C_p^{p_{\max}} \lambda^p (1-\lambda)^{p_{\max}-p} \mathbf{1}_{\{0,1,\ldots,p_{\max}\}}(p)$$

et

$$\pi(q;\mu) = C_q^{q_{\max}} \mu^q (1-\mu)^{q_{\max}-q} \mathbf{1}_{\{0,1,\ldots,q_{\max}\}}(q)$$

où $C_a^b = \frac{b!}{(b-a)!a!}$. Pour p et q fixés, le paramètre $r^{(p)}$ (resp. $\rho^{(q)}$) suit la distribution uniforme sur l'intervalle $(-1,1)^p$ (resp. $(-1,1)^q$). Enfin, pour la variance du bruit σ^2 nous utilisons la distribution inverse gamma de paramètre $\frac{\alpha}{2}$ et $\frac{\beta}{2}$ $(\alpha, \beta > 0)$:

$$\pi(\sigma^2;\alpha,\beta) = \frac{\frac{\beta}{2}^{\frac{\alpha}{2}}}{\Gamma(\frac{\alpha}{2})} (\sigma^2)^{-(1+\frac{\alpha}{2})} \exp -\frac{\frac{\beta}{2}}{\sigma^2}.$$

Nous avons ainsi introduit de façon artificielle le vecteur d'hyperparamètre $\mathcal{H}_2 = (\lambda, \mu, \alpha, \beta)$. Il apparaît clairement au cours des simulations que le choix des hyperparamètres est un élément fondamental du fonctionnement des algorithmes. Pour diminuer l'impact de ce choix et donc pour augmenter la robustesse de la méthode, nous mettons des lois de probabilités sur ces hyperparamètres. Pour λ et μ nous prenons la loi uniforme sur $[0,1]$, nous fixons $\alpha = 2$ puis nous choisissons la distribution hyper-à priori de Jeffrey pour l' hyperparamètre β. Ainsi, la structure hiérarchique se traduit dans la formule

$$\begin{aligned} &\pi(\mathcal{H}_1, \mathcal{H}_2) \\ &= \pi(p \mid \lambda)\, \pi(q \mid \mu)\, \pi(r^{(p)} \mid p)\, \pi(\rho^{(q)} \mid q)\, \pi(\sigma^2 \mid \alpha, \beta)\, \pi(\lambda)\, \pi(\mu)\, \pi(\beta). \end{aligned} \tag{17}$$

Pour la distribution jointe, on a

$$\pi(\mathcal{H}_1, \mathcal{H}_2, \mathbf{x}_n) = la(\mathbf{x}_n \mid \mathcal{H}_1)\, \pi(\mathcal{H}_1, \mathcal{H}_2).$$

où $la(\mathbf{x}_n \mid \mathcal{H}_1)$ est la vraisemblance approchée des observations.

L'estimation bayésienne de $\mathcal{H}_1$(et de $\mathcal{H}_2$) est basée sur la distribution à posteriori suivante qui est obtenue par le théorème de Bayes:

$$\underbrace{\pi(\mathcal{H}_1, \mathcal{H}_2 \mid \mathbf{x}_n)}_{\text{à posteriori}} \propto \underbrace{la(\mathbf{x}_n \mid \mathcal{H}_1)}_{\text{vraisemblance approchée}} \underbrace{\pi(\mathcal{H}_1, \mathcal{H}_2)}_{\text{à priori}}, \tag{18}$$

qui se décompose comme le produit de la fonction de vraisemblance (11) et la distribution à priori (17).

Comme les paramètres d'intérêt sont $p, q, r^{(p)}$ et $\rho^{(q)}$, on marginalise en intégrant par rapport à σ^2.

$$\pi(\mathcal{H}_1', \mathcal{H}_2 \mid \mathbf{x}_n) = \int_{\mathbb{R}^+} \pi(\mathcal{H}_1, \mathcal{H}_2 \mid \mathbf{x}_n)\, d\sigma^2$$

avec $\mathcal{H}_1' = \left(p, q, r^{(p)}, \rho^{(q)}\right)$. En posant

$$\begin{aligned} \nu &= v(\alpha) = \tfrac{\alpha}{2} + \frac{n - p_{\max}}{2} \\ w &= w(\beta, p, q, r^{(p)}, \rho^{(q)}) = \tfrac{\beta}{2} + \frac{1}{2} \sum_{t=p_{\max}+1}^{n} e^2(t, p, q, F^{-1}(r^{(p)}), G^{-1}(\rho^{(q)})) \end{aligned} \tag{19}$$

où les fonctions $e(.)$, $F(.)$, et $G(.)$ sont définiés dans (7), (8), et (9) et en utilisant

$$\int_{\mathbb{R}^+} (\sigma^2)^{-(1+v)} \exp -\frac{w}{\sigma^2} d\sigma^2 = \frac{\Gamma(v)}{w^v}$$

on obtient

$$\pi(p, q, r^{(p)}, \rho^{(q)}, \mathcal{H}_2 \mid \mathbf{x}_n) \propto C_p^{p_{\max}} \lambda^p (1-\lambda)^{p_{\max}-p} C_q^{q_{\max}} \mu^q (1-\mu)^{q_{\max}-q} (\frac{1}{2})^{p+q} \frac{\frac{\beta}{2}^{\frac{\alpha}{2}}}{\Gamma(\frac{\alpha}{2})} \frac{1}{\beta} \frac{\Gamma(\nu)}{w^\nu} \tag{20}$$

4.2 MCMC à sauts réversibles pour le modèle ARMA

Rappelons tout d'abord l'idée générale de l'algorithme de Metropolis-Hastings ([20], [13]) appelé maitenant simulation de type Monte Carlo par Chaîne de Markov (MCMC).

Soit π une mesure de probabilité sur un ensemble quelconque E. Il arrive fréquemment que π ne soit connue qu'à une constante multiplicative près (mesure de Gibbs sur un ensemble de configuration de très grande taille, par exemple) ou bien que π soit d'une complexité telle, qu'elle ne puisse se simuler facilement à l'aide des lois usuelles (mesure à postériori dans une approche bayésienne d'un modèle complexe, comme dans notre étude). Les moyens de calculs actuels ont remis à l'honneur l'idée de Metropolis-Hastings qui consiste à faire tourner une chaîne de Markov admettant comme loi stationnaire, la mesure π. La construction d'un noyau de transition d'une telle chaîne est remarquablement simple et peut s'obtenir de la façon suivante:

1. on choisit un *noyau auxiliaire de transition* quelconque Q sur E
2. à partir du point courant ξ, on génére un point $\boldsymbol{\xi}^*$ selon un noyau $Q(\boldsymbol{\xi}^* \mid \boldsymbol{\xi})$
3. on accepte ce point $\boldsymbol{\xi}^*$ avec une probabilité

$$\rho(\boldsymbol{\xi}, \boldsymbol{\xi}^*) = \min \left\{ 1, \frac{\pi(\boldsymbol{\xi}^*)}{\pi(\boldsymbol{\xi})} \frac{Q(\boldsymbol{\xi} \mid \boldsymbol{\xi}^*)}{Q(\boldsymbol{\xi}^* \mid \boldsymbol{\xi})} \right\} \tag{21}$$

On montre alors que le noyau K résultant de ces deux opérations (proposition et acceptation/rejet) vérifie les équations de réversibilité

$$\iint_{A \times B} \pi(\boldsymbol{\xi}) \, \mathbf{K}(\boldsymbol{\xi}^* \mid \boldsymbol{\xi}) \, d\boldsymbol{\xi} d\boldsymbol{\xi}^* = \iint_{B \times A} \pi(\boldsymbol{\xi}^*) \, \mathbf{K}(\boldsymbol{\xi} \mid \boldsymbol{\xi}^*) \, d\boldsymbol{\xi}^* d\boldsymbol{\xi}$$

et que sous conditions d'*irréductibilité* et d'*apériodicité* de la chaîne ainsi construite, cette *mesure invariante* est aussi la *loi stationnaire* de la chaîne. Alors, après un temps de chauffage, on peut considérer que les sorties de l'algorithme sont distribuées selon π.

Cette méthode très générale, présente une difficulté d'application lorque l'ensemble E est constitué de plusieurs sous espaces de *dimension différentes.* Supposons pour simplifier que

$$E = \mathbb{R}^{k_1} \cup \mathbb{R}^{k_2} \quad \text{avec} \quad k_1 < k_2$$

Pour définir le noyau auxiliaire Q, il faut définir un mécanisme de saut de $\mathbb{R}^{k_1}$ vers $\mathbb{R}^{k_2}$, puis de $\mathbb{R}^{k_2}$ vers $\mathbb{R}^{k_1}$, qui permettent à la chaîne d'être irréductible. P. Green a proposé une solution dans son article

de Biometrika ([12]). Pour résumer, à partir du point $\boldsymbol{\xi}$ de $\mathbb{R}^{k_1}$, on génére de façon *aléatoire* $k_2 - k_1$ composantes qui permettent d'obtenir un point de $\mathbb{R}^{k_2}$, on définit une transformation *déterministe* dans $\mathbb{R}^{k_2}$ dans lui-même ce qui donne $\boldsymbol{\xi}^*$ et enfin une transformation (généralement) déterministe de $\mathbb{R}^{k_2}$ vers $\mathbb{R}^{k_1}$ qui rende le saut de $\boldsymbol{\xi}^*$ vers $\boldsymbol{\xi}$ possible (irréductibilité). Ceci permet de définir le noyau Q, la procédure d'acceptation/rejet est alors du même type, la formule (21) faisant apparaître en plus le jacobien de la transformation faisant passer de $\boldsymbol{\xi}$ à $\boldsymbol{\xi}^*$

$$\rho(\boldsymbol{\xi}, \boldsymbol{\xi}^*) = \min \left\{ 1, \frac{\pi(\boldsymbol{\xi}^*)}{\pi(\boldsymbol{\xi})} \frac{Q(\boldsymbol{\xi} \mid \boldsymbol{\xi}^*)}{Q(\boldsymbol{\xi}^* \mid \boldsymbol{\xi})} \left| J\left(\boldsymbol{\xi}, \boldsymbol{\xi}^*\right) \right| \right\} \tag{22}$$

Dans le problème que nous étudions, le recours à la reparamétrisation (cf. § 2.4) nous fournit comme espace d'états de la chaîne, l'ensemble

$$E = \bigcup_{\substack{p=0,\dots,p_{\max} \\ q=0,\dots,q_{\max}}}]-1,+1^p[\times]-1,+1[^q$$

Nous détaillerons les sauts dans le paragraphe suivant. En résumé, les parties AR et MA seront modifiés de façon indépendante, les sauts sur les ordres seront du type $p \to p+1$ appelé *naissance d'un ordre* dans la suite, $p \to p-1$ *décès d'un ordre,* mais aussi $p \to p$ avec changement des paramètres (même chose sur q).

5 Description de l'Algorithme

Rappelons que nous souhaitons simuler le choix aléatoire de point selon la loi

$$\pi(\mathcal{H}_1^{'}, \mathcal{H}_2 \mid \mathbf{x}_n)$$

où $\mathcal{H}_1^{'}$ est le vecteur de paramètres $\left(p, q, r^{(p)}, \rho^{(q)}\right)$ et $\mathcal{H}_2$ le vecteur des hyperparamètre $(\lambda, \mu, \alpha, \beta)$. Nous adoptons un *algorithme de Gibbs hybride* (cf.[24] p. 208.) qui consiste, à partir du point courant à

1. **Etape 1.** Simuler selon $\pi\left(\mathcal{H}_2 \mid \mathcal{H}_1^{'}, \mathbf{x}_n\right)$
2. **Etape 2**. Simuler selon $\pi\left(\mathcal{H}_1^{'} \mid \mathcal{H}_2, \mathbf{x}_n\right)$

la simulation selon $\pi(\mathcal{H}_1^{'} \mid \mathcal{H}_2, \mathbf{x}_n)$ se faisant grâce à une étape de l'algotithme MCMC à sauts réversibles. L'Etape 1 a pour but d'actualiser les hyperparamètres au cours de la simulation.

5.1 Actualisation des hyperparamètres

Grâce à la structure produit de la loi sur les hyperparamètres, il est facile de voir que pour λ et μ, on a

$$\begin{aligned} & \pi\left(\lambda \mid \mathcal{H}_1^{'}, \mathbf{x}_n\right) = \pi\left(\lambda \mid p\right) = \mathcal{B}\left(p+1, p_{\max} - p + 1\right) \\ \text{et} \quad & \pi\left(\mu \mid \mathcal{H}_1^{'}, \mathbf{x}_n\right) = \pi\left(\mu \mid q\right) = \mathcal{B}\left(q+1, q_{\max} - q + 1\right) \end{aligned}$$

où $\mathcal{B}\left(a, b\right)$ désigne la loi Bêta de paramètres a et b.

Pour l'actualisation du paramètre β, on doit tout d'abord actualiser σ^2, ce qui donne

1. Simuler σ^2 selon $\pi\left(\sigma^2 \mid \beta, \mathcal{H}_1^{'}, \mathbf{x}_n\right)$
2. Simuler β selon $\pi\left(\beta \mid \sigma^2, \mathcal{H}_1^{'}, \mathbf{x}_n\right)$.

Il est facile de voir, par conjuguaison que

$$\pi\left(\beta \mid \sigma^2\right) \propto \beta^{\frac{\alpha}{2}-1} \exp - \left[\frac{1}{2\sigma^2}\beta\right] \tag{23}$$

c'est à dire que, conditionnellement à σ^2 ($\alpha = 2$ fixé), β suit une loi $\mathcal{G}(\frac{\alpha}{2}, \frac{1}{2\sigma^2})$. Ce choix est le même que dans [21] . D'autre part,

$$\pi\left(\sigma^2 \mid \mathcal{H}_1^{'}, \mathbf{x}_n\right) \propto la(\mathbf{x}_n \mid \mathcal{H}_1^{'})\ \pi(\sigma^2 \mid \beta)\ \pi(\beta) \propto (\sigma^2)^{-(\upsilon+1)} \exp - \frac{\omega}{\sigma^2} \tag{24}$$

où υ et w sont définis dans (19). Il s'agit de la loi inverse Gamma de paramètre (υ, w).

En résumé, si à l'étape j de l'algorithme, la valeur des paramètres est $\left(p, q, r^{(p)}, \rho^{(q)}\right)$ et des hyperparamètres $(\lambda, \mu, \alpha, \beta)$ l'étape d'actualisation des hyperparamètres est

1. Génération de $\lambda^{(j+1)}$ selon $\mathcal{B}\left(p+1, p_{\max} - p + 1\right)$
2. Génération de $\mu^{(j+1)}$ selon $\mathcal{B}\left(q+1, q_{\max} - q + 1\right)$
3. Génération de σ^2 selon (23)
4. Génération de $\beta^{(j)}$, voir équation (24)

5.2 Sauts sur les paramètres

Pour simuler selon $\pi\left(\mathcal{H}_1^{'} \mid \mathcal{H}_2, \mathbf{x}_n\right)$ nous utilisons une étape de simulation de type MCMC à sauts réversibles, qui se décompose de la façon suivante.

5.2.1 Choix du type de sauts

Si (p, q) est le point courant pour les ordres, on notera

$$\begin{cases} \eta_p^{AR} & \text{la probabilité de sauts de} & p \curvearrowright p+1 \\ \eta_q^{MA} & \cdots & q \curvearrowright q+1 \\ \delta_p^{AR} & \cdots & p \curvearrowright p-1 \\ \delta_q^{MA} & \cdots & q \curvearrowright q-1 \\ \varsigma_p^{AR} & \cdots & p \curvearrowright p \\ \varsigma_q^{MA} & \cdots & q \curvearrowright q \end{cases}$$

Pour chaque composante, on choisira la probabilité uniforme sur l'ensemble des posibilités de sauts, donc généralement une probabilité de $\frac{1}{3}$ sauf pour les valeurs extrêmes de p et q. Ceci donne l'algorithme

suivant à partir du point (p, q)

si $u_1 \sim \mathcal{U}(0,1) \leq \eta^{AR}$	alors	$p \curvearrowright p+1$
si $u_1 \sim \mathcal{U}(0,1) \leq \eta_p^{AR} + \delta_p^{AR}$	$\cdots$	$p \curvearrowright p-1$
sinon	$\cdots$	$p \curvearrowright p$
si $u_2 \sim \mathcal{U}(0,1) \leq \eta_q^{MA}$	$\cdots$	$q \curvearrowright q+1$
si $u_2 \sim \mathcal{U}(0,1) \leq \eta_q^{MA} + \delta_q^{MA}$	$\cdots$	$q \curvearrowright q-1$
sinon	$\cdots$	$q \curvearrowright q$

Nous décrivons dans ce qui suit chaque type de sauts et nous calculons la probabilité d'acceptation/rejet associée.

5.2.2 Naissance/décès d'une ordre

Nous raisonnons sur la partie AR. La structure particulière de l'espace des nouveaux paramètres de l'ARMA autorise des sauts très simples. Soit p la valeur courante de l'ordre de l'ARMA, $r^{(p)} = (r_1, ..., r_p)$ le vecteur de ses coefficients et supposons que nous effectuions un saut $p \curvearrowright p+1$. On tire alors une variable aléatoire u selon *la loi triangulaire centrée en 0* [4]

$$g(u) = \begin{cases} u+1, & -1 < u < 0 \\ 1-u, & 0 < u < 1 \end{cases} . \tag{25}$$

On complète le vecteur $r^{(p)}$ par cette composante aléatoire et le nouveau vecteur de coefficients proposé sera ce point, soit

$$r^{(p+1)} = (r_1, ..., r_{p,} u)$$

Notons que, si l'on traduit cette transformation sur les anciens paramètres, ceci revient à un changement total de la valeur de ceux-ci. Il est clair que le Jacobien de la transformation vaut 1.

Inversement, le saut de type $p+1 \curvearrowright p$ consistera à supprimer la dernière composante dans $r^{(p+1)} = (r_1, ..., r_p, r_{p+1})$ soit à proposer le point $r^{(p)} = (r_1, ..., r_p)$. La probabilité d'acceptation/rejet se calcule conjointement pour ces deux types de sauts

$$\alpha_{\mathcal{N}} = \min\{1,\, r_{\mathcal{N}}\} \quad \text{et} \quad \alpha_{\mathcal{D}} = \min\{1,\, r_{\mathcal{N}}^{-1}\} \tag{26}$$

avec

$$r_{\mathcal{N}} = \frac{\pi(p+1, q, r^{(p+1)}, \rho^{(q)} \mid \mathcal{H}_2, \mathbf{x}_n)}{\pi(p, q, r^{(p)}, \rho^{(q)} \mid \mathcal{H}_2, \mathbf{x}_n)} \frac{Q(p, r^{(p)} \mid p+1, r^{(p+1)})}{Q(p+1, r^{(p+1)} \mid p, r^{(p)})}$$

On a

$$\begin{cases} Q(p, r^{(p)} \mid p+1, r^{(p+1)}) = \delta_{p+1}^{AR} \\ Q(p+1, r^{(p+1)} \mid p, r^{(p)}) = \eta_p^{AR}\, g(r_{p+1}) \end{cases}$$

[4] Nous n'avons pas, à dessein, repri la loi uniforme sur $]-1, +1[$ qui est la loi à priori sur r_{p+1}.

Finalement, en utilisant (20) on obtient

$$r_{\mathcal{N}} = \frac{w(\beta, p+1, q, r^{(p+1)}, \rho^{(q)})^{-\upsilon(\alpha)}}{w(\beta, p, q, r^{(p)}, \rho^{(q)})^{-\upsilon(\alpha)}} \frac{p_{\max} - p}{p+1} \frac{1-\lambda}{\lambda} \frac{1}{2} \frac{\eta_{p+1}^{AR}}{\delta_p^{AR}} \frac{1}{g(r_{p+1})}$$

où les quantités w et υ sont données dans (19) .

Si une naissance a été choisie, la partie correspondante de l'algorithme est

---

1. Générer $u \sim q(.)$, voir l'équation (25).
2. Evaluer $\alpha_{\mathcal{N}}$, voir l'équation (26).
3. Si $v \sim \mathcal{U}(0,1) \leq \alpha_{\mathcal{N}}$ nouveau coefficient : $r^{(p+1)} = (r_1, ..., r_p, u)$ sinon le coefficient de la part AR reste $r^{(p)}$.

Si un décès est choisi

---

1. Suprimer r_{p+1}.
2. Evaluer $\alpha_{\mathcal{D}}$ voir l'équation (26).
3. Si $v \sim \mathcal{U}(0,1) \leq \alpha_{\mathcal{D}}$, nouveau coefficient $r^{(p)}$de la part AR sinon le coefficient de la part AR reste $r^{(p+1)}$.

5.2.3 Changement des coefficients:

Supposons maintenant que le saut choisi sur la partie AR soit $p \curvearrowright p$ donc pas de changement de l'ordre de l'ARMA mais seulement changement des coefficients. Si $r^{(p)} = (r_1, ..., r_p)$ est le vecteur de ces coefficients, on modifie ce vecteur composante par composante, faisant ainsi du *Metropolis Hastings composante pas composante* ([11] p. 10). On actualise successsivement chaque composante en commencant par r_1, r_2 ... Supposons que $r_1, \ldots, r_{i-1}$ aient été actualisés et notons $r_1^*, \ldots, r_{i-1}^*$ les points obtenus. On génére un point u_i selon la *loi triangulaire centrée* en r_i

$$q(u_i \mid r_i) = \begin{cases} \dfrac{(u_i+1)}{(r_i+1)}, & -1 < u_i < r_i \\ \dfrac{(1-u_i)}{(1-r_i)}, & r_i < u_i < 1 \end{cases} . \tag{27}$$

Cette loi triangulaire a l'avantage de tenir compte du point courant et de charger plus fortement le voisinage de ce point, donc de générer de petits déplacements. Un autre choix envisageable, et qui a été utilisé avec succès dans les simulations, est de modifier globalement le vecteur $r^{(p)} = (r_1, ..., r_p)$ avec un tirage selon une loi proportionnelle à

$$J_{F^{-1}}(r^{(p)}) = \prod_{k=2}^{p} (1+r_k)^{[k/2]} (1-r_k)^{[(k-1)/2]}$$

c'est à dire un produit de loi Bêta, de paramètres dépendant de l'indice du coefficient. Comme cette loi est proportionnelle au Jacobien de la transformation F liant les anciens et nouveaux paramètres, ceci revient à faire un tirage selon la *loi uniforme* dans le domaine de causalité $\mathcal{C}_p$ des anciens paramètres.

En notant $r^{*(i)} = (r_1^*, \ldots, r_{i-1}^*, u_i, r_{i+1}, \ldots, r_p)$ et $\widetilde{r^{*(i)}} = (r_1^*, \ldots, r_{i-1}^*, r_i, r_{i+1}, \ldots, r_p)$ la probabilité d'acceptation/rejet associée est

$$\alpha_{\mathcal{C}} = \min\{1,\, r_{\mathcal{C}}\} \tag{28}$$

où

$$r_{\mathcal{C}} = \frac{\pi(p, q, r^{*^{(i)}}, \rho^{(q)}, \mathcal{H}_2 \mid \mathbf{x}_n)}{\pi(p, q, \widetilde{r^{*(i)}}, \rho^{(q)}, \mathcal{H}_2 \mid \mathbf{x}_n)} \frac{Q(p, \widetilde{r^{*(i)}} \mid p, r^{*^{(i)}})}{Q(p, r^{*^{(i)}} \mid p, \widetilde{r^{*(i)}})}$$

où

$$\frac{Q(p, \widetilde{r^{*(i)}} \mid p, r^{*^{(i)}})}{Q(p, r^{*^{(i)}} \mid p, \widetilde{r^{*(i)}})} = \frac{1 + r_i}{1 + u_i} \frac{1 - r_i}{1 - u_i}$$

Grâce à l'équation (20) on obtient:

$$r_{\mathcal{C}} = \frac{w(\beta, p, q, r^{*(p)}, \rho^{(q)})^{-\upsilon(\alpha)}}{w(\beta, p, q, r^{(p)}, \rho^{(q)})^{-\upsilon(\alpha)}} \frac{1 + r_i}{1 + u_i} \frac{1 - r_i}{1 - u_i},$$

avec $\upsilon(.)$ et $w(.)$ sont définis dans (19). L'algorithme est la suivante:

Pour $i = 1, ..., p$

1. Générer $u_i \sim q(.)$, voir l'équation (27).
2. Evaluer $\alpha_{\mathcal{C}}$, voir l'équation (28).
3. Si $v \sim \mathcal{U}(0,1) \leq \alpha_{\mathcal{C}}$ nouveau coefficient $r_i^* = u_i$
   sinon $r_i^* = r_i$

Il est facile de vérifier que l'ensemble de ces possibilités de sauts rend la chaîne de Markov irréductible dans

$$E = \cup_{\substack{p=0,\ldots,p_{\max} \\ q=0,\ldots,q_{\max}}}]-1,+1^p[\times]-1,+1[^q$$

et apériodique.

5.3 Estimation

Dans l'inférence bayésienne, les estimations sont entièrement basées sur la distribution à postériori. L'algorithme de MCMC à sauts réversibles détaillé dans la section précédente génére un échantillon $(\boldsymbol{\varsigma}^{(i)}; i = 1, ..., M)$ distribué selon la distribution à posteriori $\pi(\mathcal{H}_1', \mathcal{H}_2 \mid \mathbf{x}_n)$. Nous avons alors développer deux types d'estimation.

Tout d'abord, à partir de cet échantillon, on détermine la *distribution Marginale A Postériori (MAP)* de p et q en traçant l' histogramme associé aux sorties de l'algorithme. En appelant $\widehat{\pi}(p = l \mid \mathbf{x}_n)$ $(l = 0, \ldots, p_{\max})$ cette loi, et une notation similaire pour q, on estime p (resp. q) par Maximum de Vraisemblance à savoir

$$\widehat{p} = \underset{p^{(i)} \in \{0,1,\ldots,p_{\max}\}}{\arg\max} \widehat{\pi}(p = l \mid \mathbf{x}_n),$$
$$\widehat{q} = \underset{q^{(m)} \in \{0,1,\ldots,q_{\max}\}}{\arg\max} \widehat{\pi}(q = m \mid \mathbf{x}_n).$$

On estime ensuite les paramètres $r^{(p)}$ et $\rho^{(q)}$ par les *Moyennes Marginales à Postériori* obtenues à partir des sorties de l'algorithme associées aux ordres $\widehat{p}$ et $\widehat{q}$.

Nous avons également utlilisé la technique du Recuit Simulé ([10], [16]) pour obtenir directement le *Maximum à Postériori* sur l'ensemble des paramètres du modèle. Rappelons que pour ce faire, il suffit d'utiliser le même noyau de transition de la chaîne de Markov, en ajoutant un schéma de descente de température.

5.4 Résutats et simulations

Pour illustrer les méthodes proposées ci-dessus, nous avons traité quatre exemples et les estimations des paramètres sont comparées à ceux du logiciel PEST proposé par Brockwell et Davis ([7]). Deux de ces exemples traitent de données réelles, le niveau du lac Huron et l'indice Dow-Jones. Les deux autres sont des données simulées. Le logiciel que nous utilisons a été écrit sous MATLAB. Dans chaque exemple nous avons pris $p_{\max} = q_{\max} = 10$

Pour l'algorithme MCMC à saut réversibles, noté simplement "Algorithme MCMC" dans la suite, nous avons fait 50.000 itérations du processus de saut, dont 20.000 itérations de chauffage. Dans sa version " Recuit Simulé", noté Algorithme RS, nous avons également fait 50.000 itérations en prenant comme température initiale $T_0 = 10$ et une réduction d'un facteur 0.99 toutes les 100 itérations.

Pour le quatrième exemple, nous avons fait tourner 50 fois l'algorithme pour étudier la robustesse de nos estimations.

5.5 Données réelles

Nous traitons des données réelles proposées par Brockwell and Davis: le niveau du lac Huron (mesuré en pieds) de 1875 à 1972 et l'évolution de l'indice Dow-Jones du 28 Août au 18 Décembre 1972.

Exemple 1 : Niveau du lac Huron

L'échantillon est composé de $n = 98$ observations et est représenté sur la figure 1.

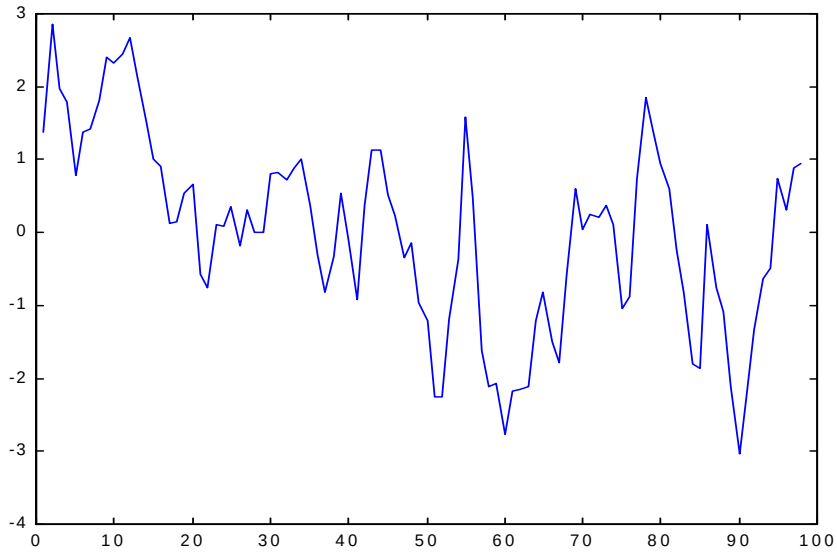


Figure 1: Niveau centré du lac Huron

Pour des processus AR "pur" ou MA "pur" Brockwell et Davis utilise la Fonction d'Autocorrélation Partielle (FACP) et la Fonction d'Autocorrélation (FAC) pour avoir une première estimation des ordres p et q respectivement. Ceci s'appuie sur le fait que si le processus AR est d'ordre p alors les Autocorrélations Partielle d'ordre k avec $k > p$ sont nulles et si le processus MA est d'ordre q les Autocorrélations d'ordre l sont nulles pour $l > q$. A partir des données, ces quantités sont estimées. Il s'en suit, qu'il est délicat déterminer le seuil à partir duquel, on peut considérer qu'elles sont nulles et cette méthode ne donne qu'une première idée d'un intervalle de variation des ordres. De plus, comme le souligne Brockwell et Davis (p. 296), l'utilisation de ces quantités devient très difficile à interpréter pour des processus ARMA. Par exemple, pour les variations du lac Huron, ces quantités sont représentées dans les Figure (2a) et (2b). Elles ne laissent guère prévoir que la modélisation retenue, par Brockwell et Davis et par notre méthode sera un ARMA$(1, 1)$.

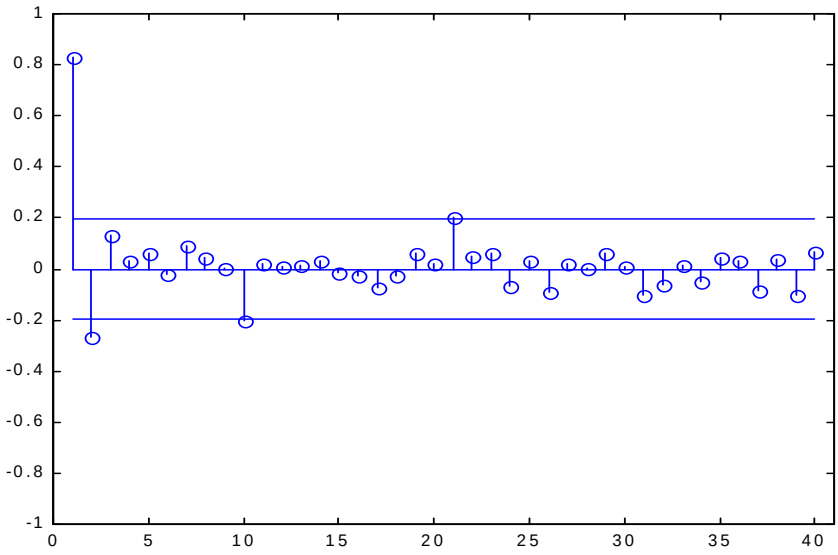


Figure 2a: Estimateurs de FACP

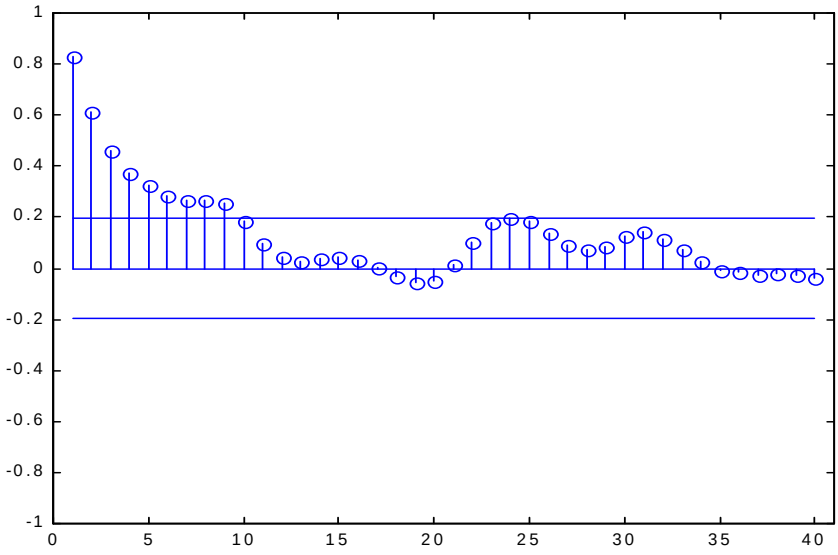


Figure 2b: Estimateurs de FAC

En fait pour les processus ARMA, Brockwell et Davis (Brockwell *et al.,* page 172) cherche à minimiser les critère AICC défini par (16) (resp. le critère AIC et le critère BIC). Diverses méthodes (cf. p 296) sont utilisées par ces auteurs pour limiter le nombre de couple (p, q) à prendre en compte dans ces calculs. Pour les variations du lac Huron, ils se limitent à $p+q \leq 6$. Pour chacun de ces critères, le mimimum est atteint pour $p = 1$ et $q = 1$. Les estimations des paramètres du processus ARMA sont alors obtenues par les méthodes classiques du maximum de vraisemblance rappelées dans la paragraphe (3.1) ([7]) et sont présentées dans le tableau 3.

Comme indiqué précédemment, notre étude est basée sur le maximum des lois marginales à postériori pour les paramètres p et q (Figure 3a et 3b). Nous trouvons, comme Brockwell et Davis, un maximum à postériori pour $p = 1$ et $q = 1$.

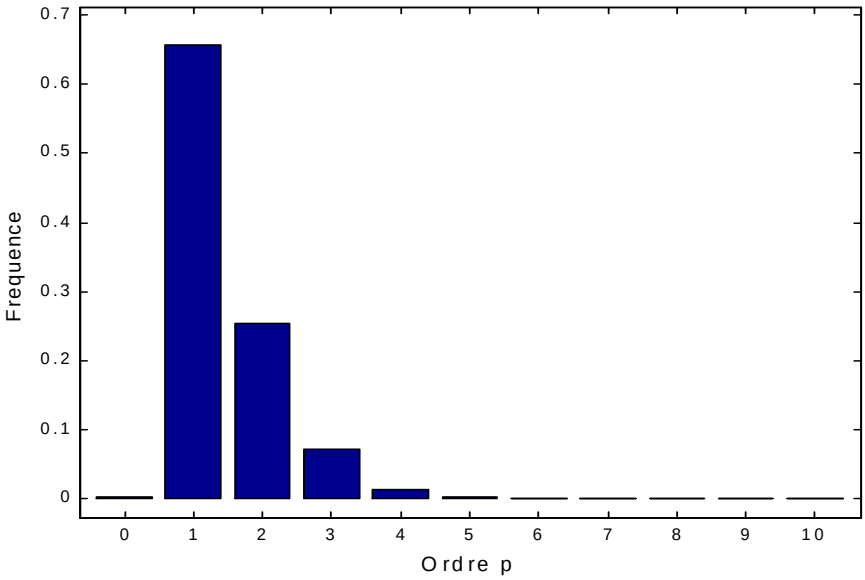


Figure 3a: Histogramme de la loi MAP de p

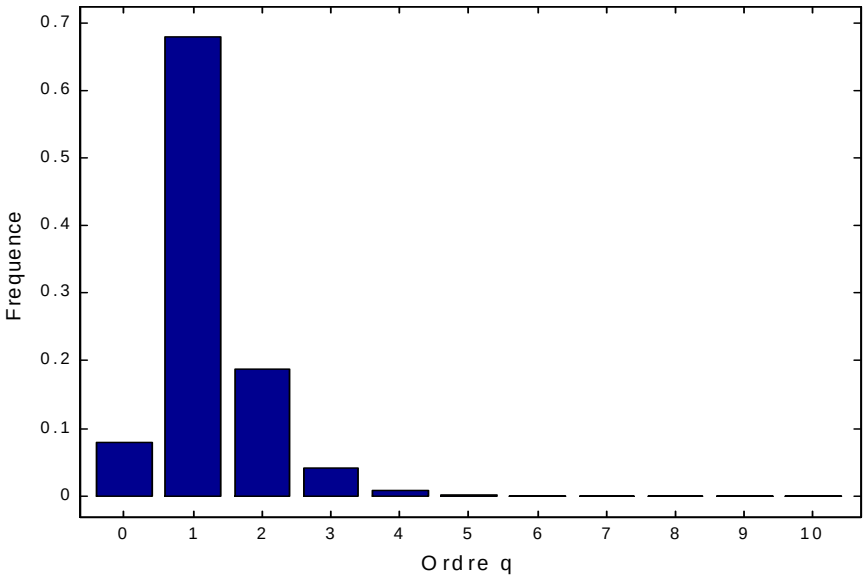


Figure 3b: Histogramme de la loi MAP de q

Pour $p = 1$ et $q = 1$ nous estimons les valeurs des paramètres par la Moyenne Marginale à Postériori associée à la sortie de l'algorithme.

L'algorithme du Recuit Simulé se stabilise également sur $p = 1$ et $q = 1$ avec des valeurs pour les paramètres, qui sont données dans le tableau de résultats suivants et comparées aux autres méthodes.

	PEST	Moy MAP	Recuit Simulé
$\phi_1^{(1)}$	-0.7453	-0.6884	-0.7326
$\theta_1^{(1)}$	0.3208	0.3627	0.3575

Tableau 1: Estimation de $\phi_1^{(1)}$ et $\theta_1^{(1)}$.

Plutôt que de se limiter à une estimation ponctuelle des paramètres de l'ARMA$(1,1)$ et pour utiliser plus complètement les sorties de l'algorithme, nous faisons figurer dans la figure (4) l'estimation par un noyau gaussien (d'écart type 0.2) des densités marginales à postériori de $\phi_1^{(1)}$ et $\theta_1^{(1)}$.

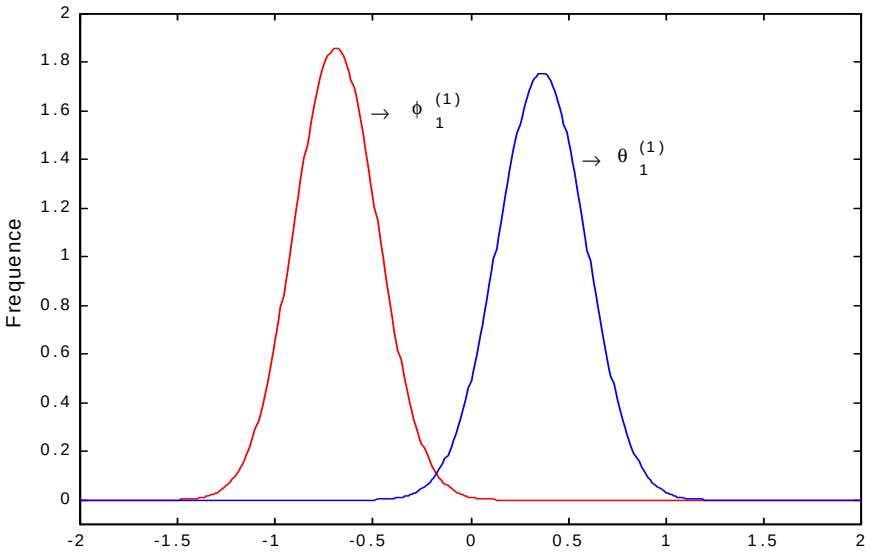


Figure 4: Estimation par noyau des lois MAP des coeficients.

Dans l'Algorithme du Recuit Simulé, on a examiné l'évolution des valeurs de p et q au cours de l'algorithme et ces tracés sont représentés figure (5).

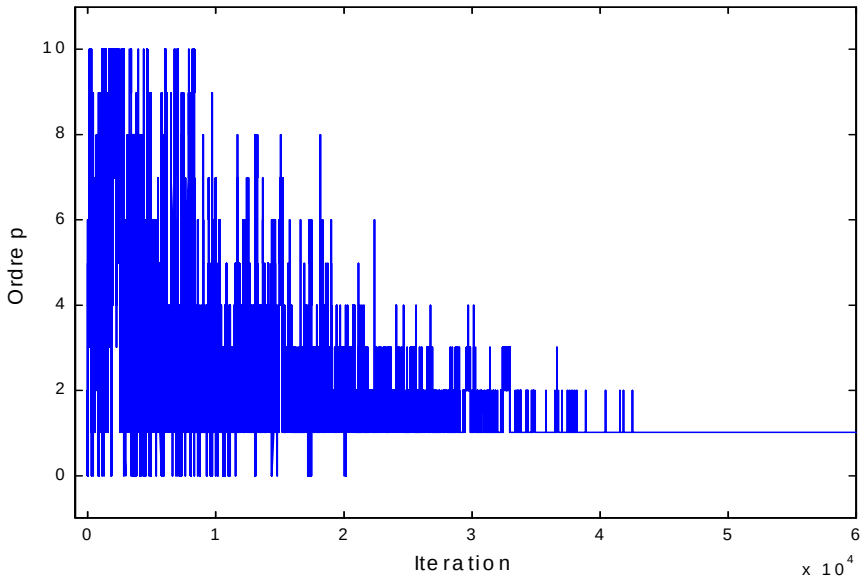


Figure 5a: Evolution de l'ordre p au cours de l'Algorithme RS

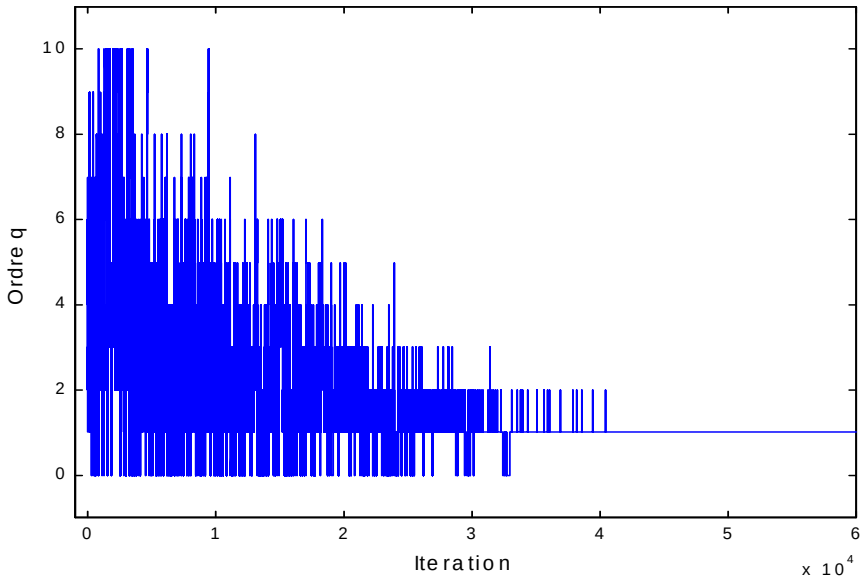


Figure 5b: Evolution de l'ordre q au cours de l'Agorithme RS

Exemple 2: **Evolution de l'Indice Dow-Jones (28 Août - 18 Décembre 1972).**

Pour ce processus, qui n'est pas stationnaire, on commence par travailler sur le *processus des différences*

$$Y_t = X_{t+1} - X_t$$

Ce processus, qui comporte $n = 77$ données est représenté sur la figure (6)

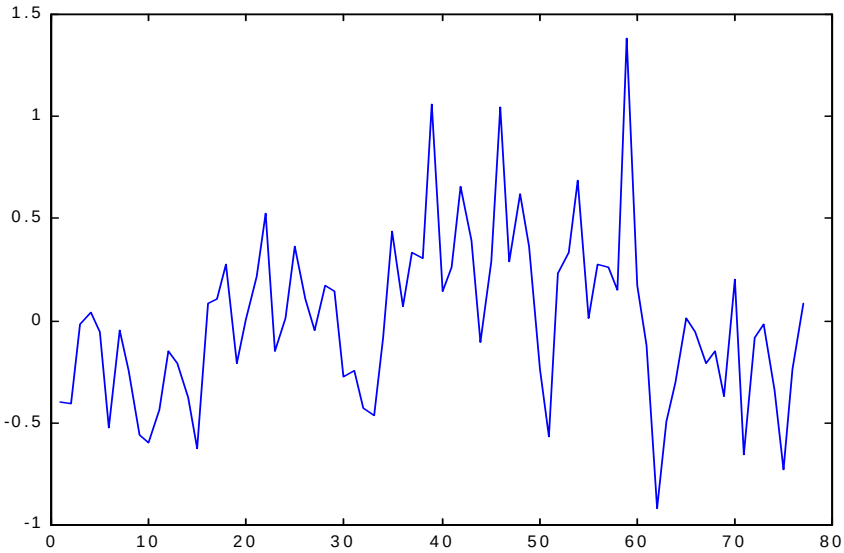


Figure 6 : Différences des Indices Dow-Jones

L'étude des critères AIIC, AIC et BIC amènene Brockwell et Davis à modéliser ces données par un processus AR(1) . Dans ce cas, l'examen de la Fonction d'Autocorrélation Partielle vient confirmer ce fait. La normalité, pour les grandes valeurs de m, de l'estimateur de la Fonction d'Autocorrélation Partielle $\widehat{r_m}$ pour un Processus AR, permet de prendre comme première estimation de p la plus petite valeur de k telle $|\widehat{r_m}| \leq 1.96\, n^{-1}$ pour $m > k$. La figure suivante donne le tracé de cette fonction, les lignes pointillées figurant la borne $\pm 1.96\, n^{-1/2}$.

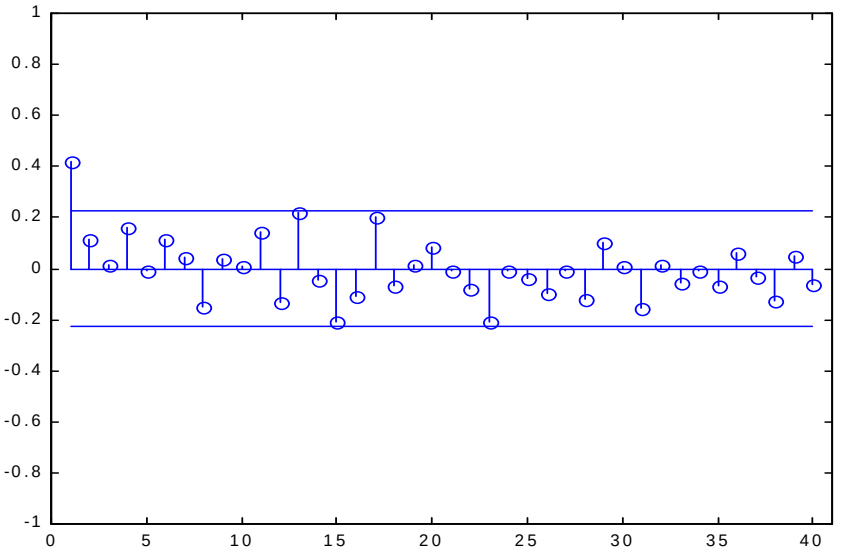


Figure 7 : Estimateurs de FACP

Comme Brockwell et Davis, l'histogramme obtenu à partir de nos simulations, nous amène à modéliser Y_t par un AR(1).

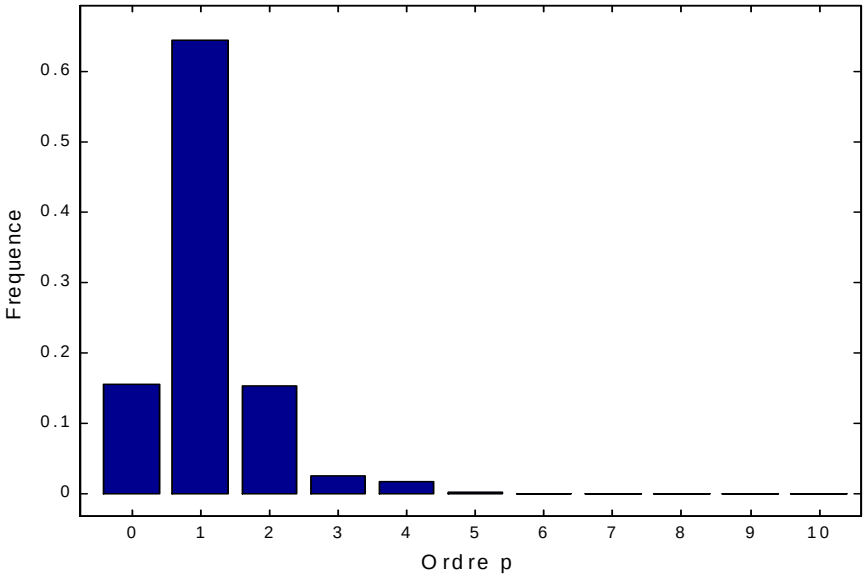


Figure 8a: Histogramme de la loi MAP de p

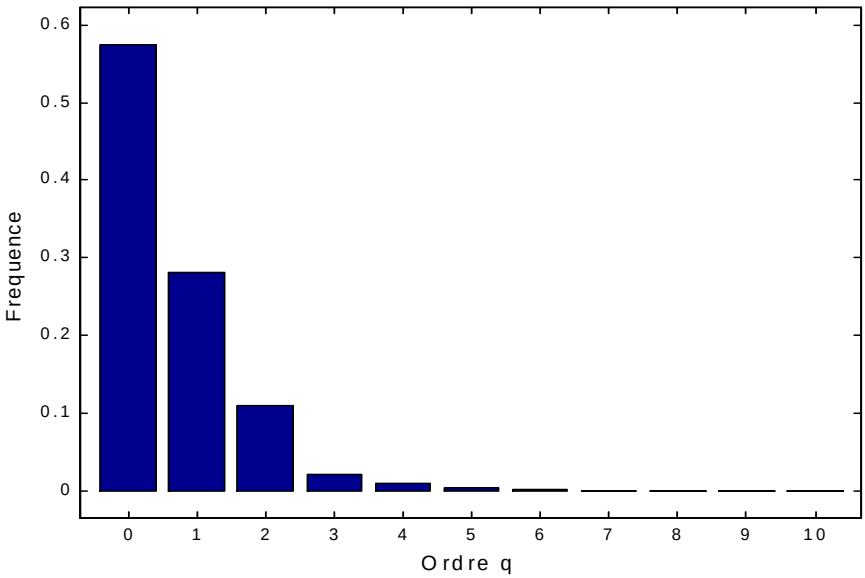


Figure 8b: Histogramme de la loi MAP de q

La version Recuit Simulé confirme ce résultat en se stabilisant sur $p = 1$ et $q = 0$. Les estimations des paramètres dans chacune des méthodes sont présentées dans le tableau suivant.

	PEST	Moy MAP	Recuit Simulé
$\phi_1^{(1)}$	-0.4471	-0.3865	-0.4630

Tableau2: Estimation de $\phi_1^{(1)}$.

La figure (9) illustre l'evolution de (p,q) au cours de l'algoritme Recuit Simulé.

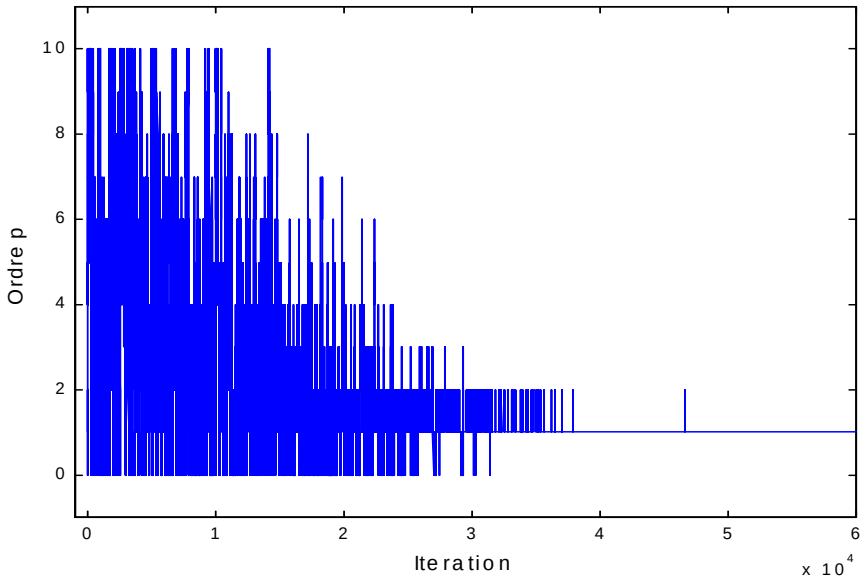


Figure 9a: Evolution de p au cours de l'Algorithme RS

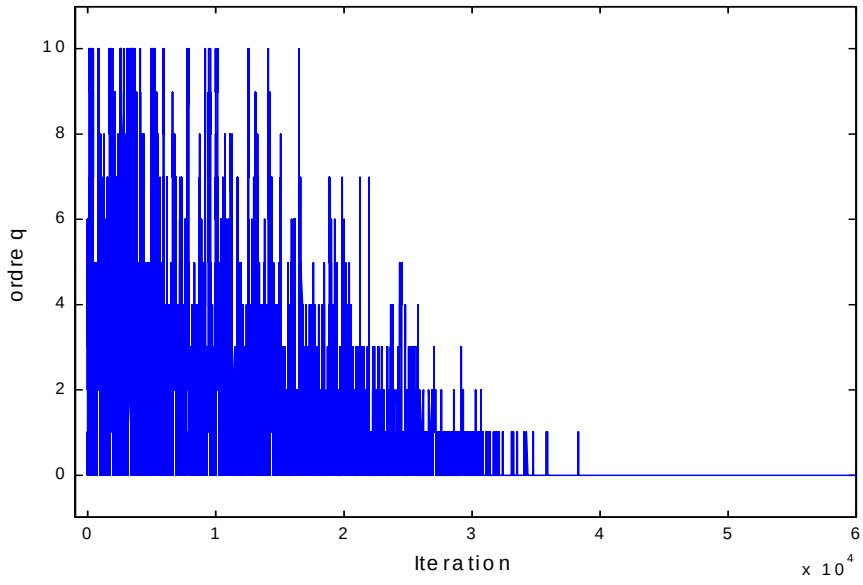


Figure 9b: Evolution de q au cours de l'Algorithme RS

L'estimation par noyau de la densité à postériori pour $\phi_1^{(1)}$ est représentée sur la figure (10).

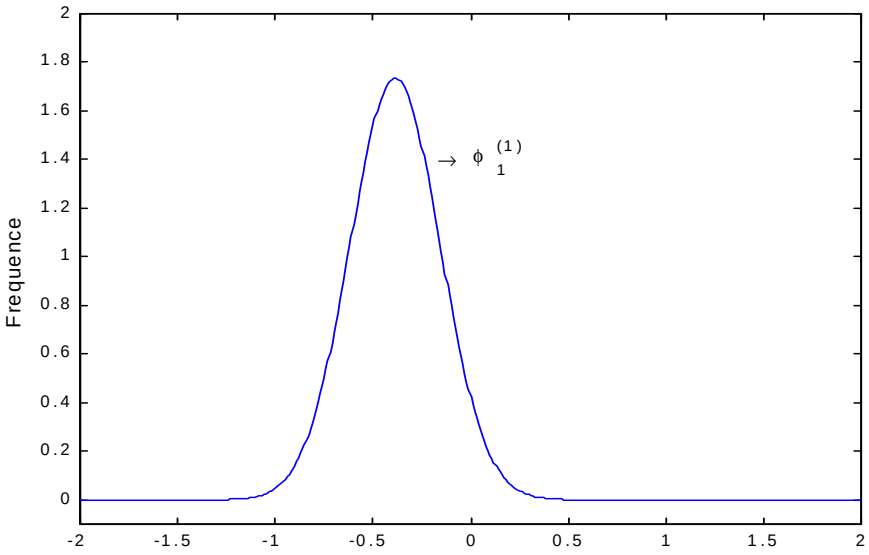


Figure 10: Estimation par noyau de la loi MAP du coefficient.

5.6 Données simulées.

Dans les exemples 3 et 4 ci-dessous, nous considérons $n = 250$ données simulées selon un processus ARMA et nous comparons les valeurs estimées par chacune des méthodes aux *vraies valeurs.*

Exemple 3: **Un Processus MA(3).**

Les paramètres du processus simulé sont résumés dans le Tableau (3).

p	q	$\phi_1^{(3)}$	$\phi_2^{(3)}$	$\phi_3^{(3)}$	σ^2
0	3	0.59	-0.3616	-0.8936	0.49

Tableau 3: Paramètres MA(3)

La figure suivante donne une réalisation du signal.

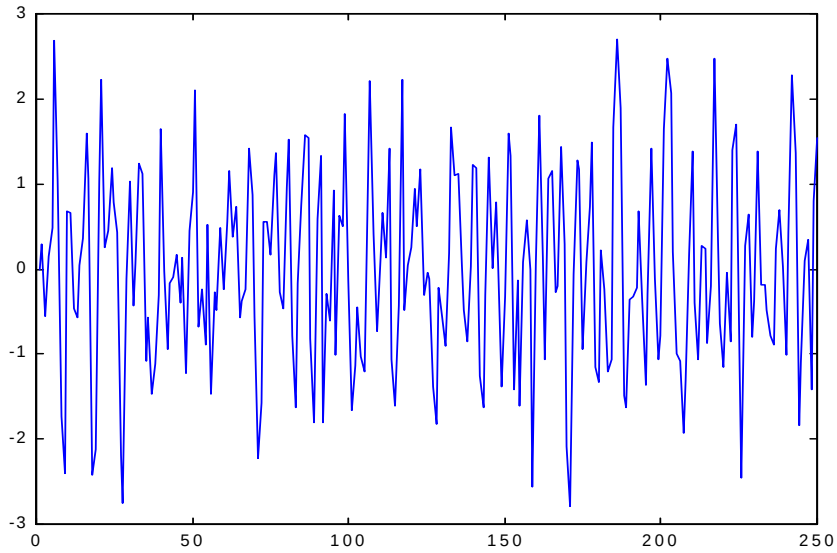


Figure 11 : Donné es MA(3)

Les critères utilisés par Brockwell et Davis donnent une modélisation par un processus MA(3) . La borne permettant de le vérifier sur l'estimation de la Fonction d'Autocorrélation est ici plus complexe (*cf. [7]* p. 247*)*. La figure suivante traduit ces résultats.

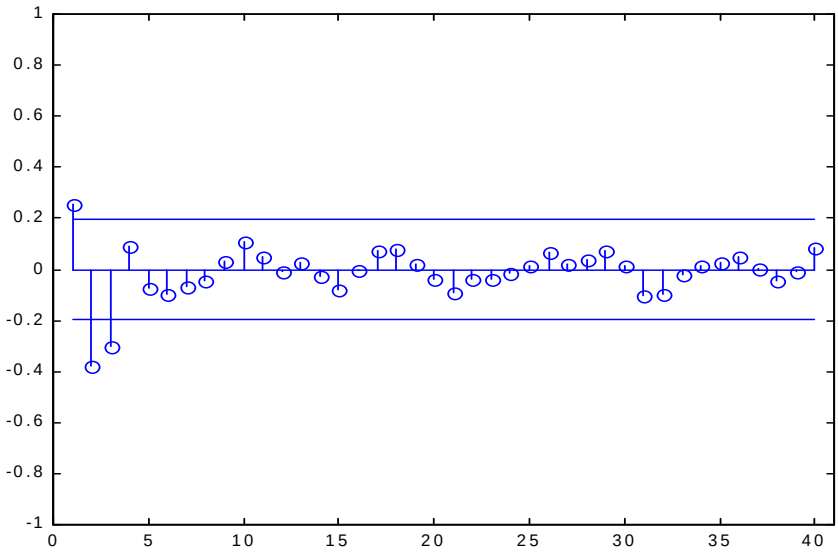


Figure 12 : Estimateurs de FAC

Nos histogrammes sur p et q sont les suivants et nous donnent $\widehat{p} = 0$ et $\widehat{q} = 3$.

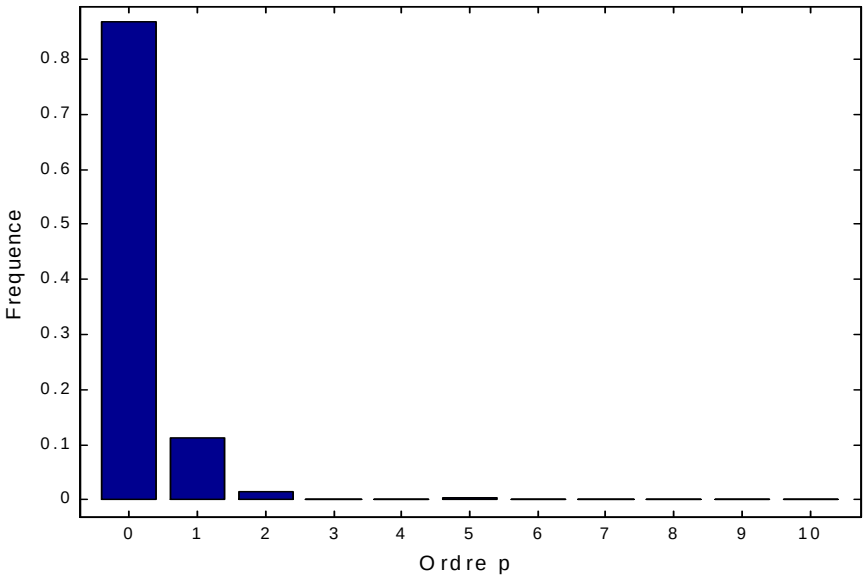
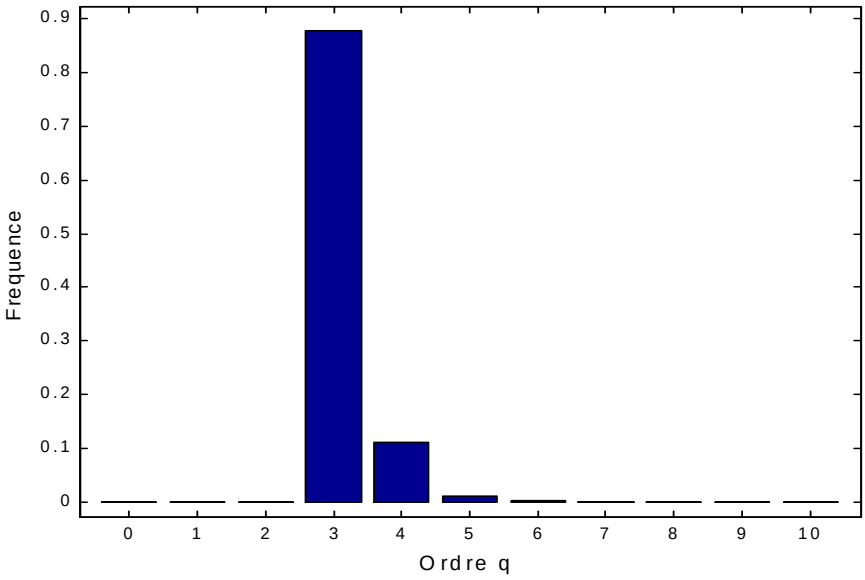


Figure 13a: Histogramme de la loi MAP de p.



Histogramme de la loi MAP de q.

Les figures qui suivent donnent l'évolution de p et q au cours la simulation de l'Algorithme RS ainsi que l'estimation par noyau des densités à postériori des trois coefficienst.

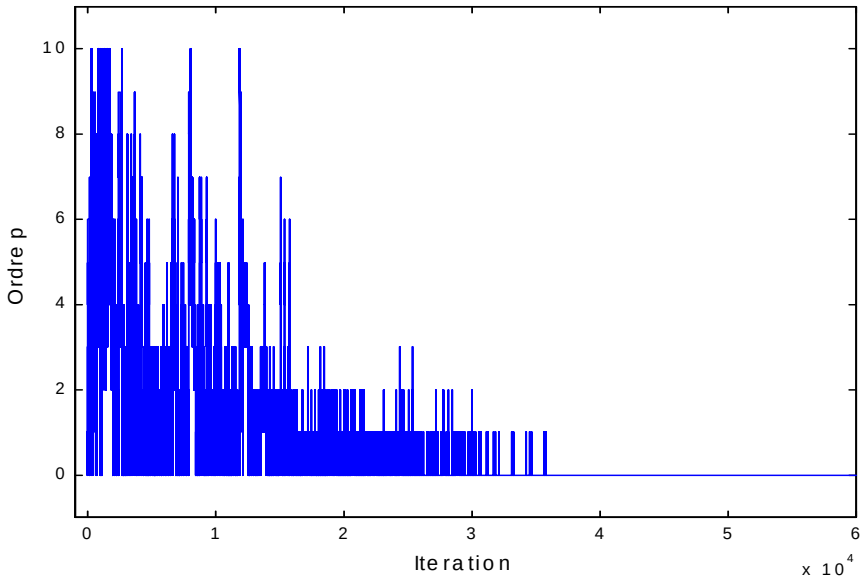


Figure 14a: Evolution de p au cours de l'Agorithme RS

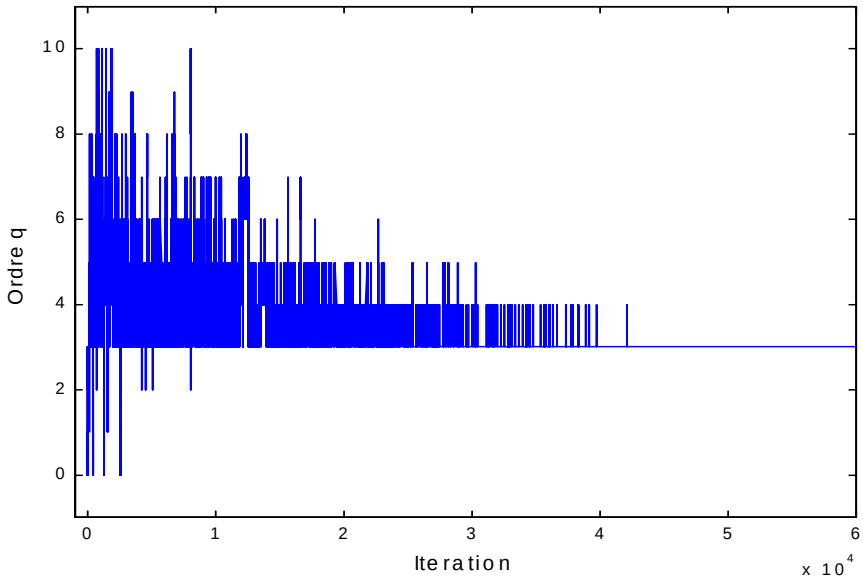


Figure 14b: Evolution de q au cours de l'Algorithme RS

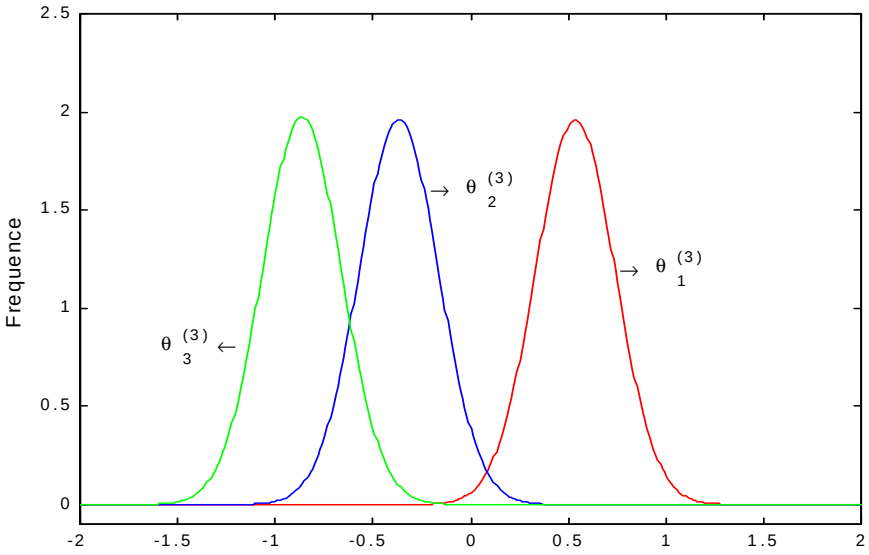


Figure 15: Estimation par noyau des lois MAP des coefficients.

Les valeurs des différentes estimations sont données dans le tableau suivant.

	Valeur vrai	PEST	Moy MAP	Recuit Simulé
$\theta_1^{(3)}$	0.5900	0.5648	0.5362	0.5575
$\theta_2^{(3)}$	-0.3616	-0.3746	-0.3676	-0.3708
$\theta_3^{(3)}$	-0.8936	-0.9045	-0.8644	-0.9059

Tableau 4: Estimations de $\theta_1^{(3)}$, $\theta_2^{(3)}$, et $\theta_3^{(3)}$.

Exemple 4: Un Processus ARMA(2,1)

Enfin un processus ARMA$(2,1)$ dont les paramètres sont donnés par le Tableau 5 est simulé. Les différents résultats sont présentés dans les figures et tableaux qui suivent.

p	$\phi_1^{(2)}$	$\phi_2^{(2)}$	q	$\theta_1^{(1)}$	σ^2
2	-1.36	0.7	1	0.7	1

Tableau 5: Paramètres ARMA(2,1)

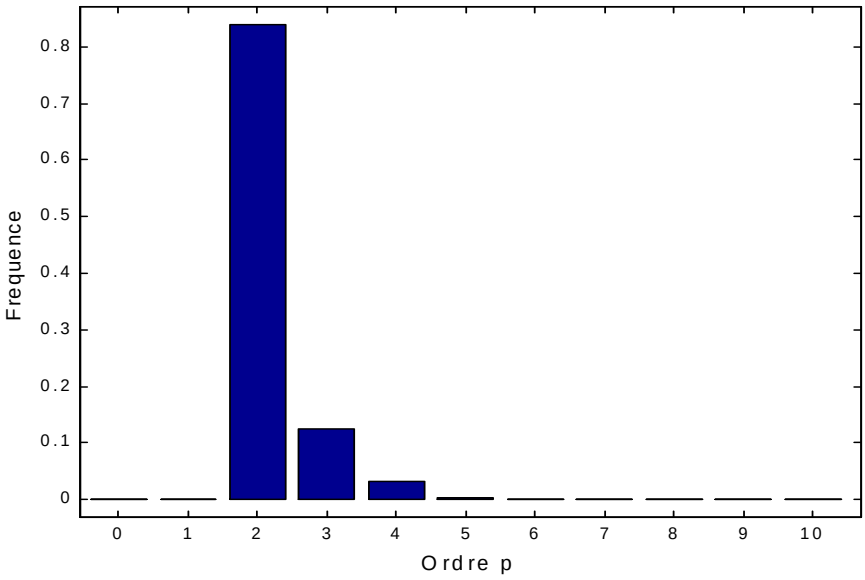


Figure 16a: Histogramme de la loi MAP de p.

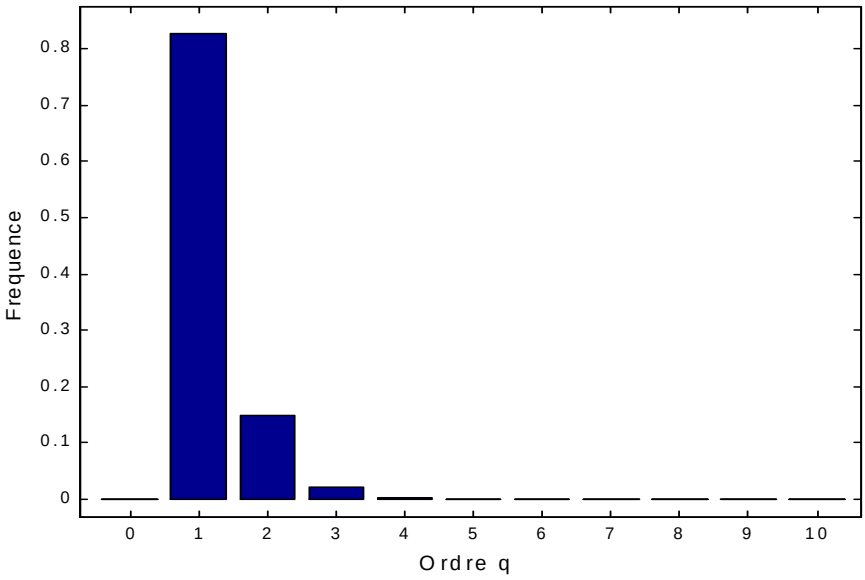


Figure 16b: Histogramme de la loi MAP de q.

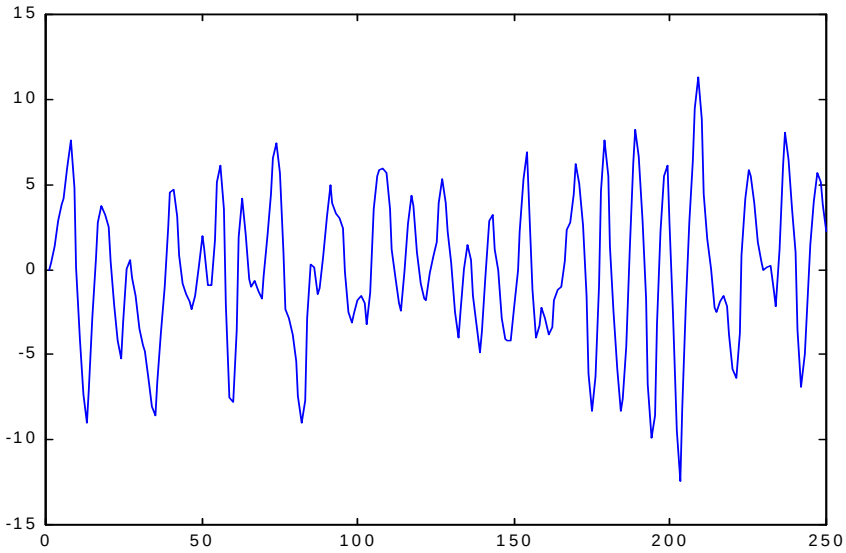


Figure 17: Observations ARMA(2,1)

	Valeur vrai	PEST	Moy MAP	Recuit Simulé
$\phi_1^{(2)}$	-1.36	-1.3941	-1.3888	-1.4091
$\phi_2^{(2)}$	0.7	0.7408	0.7355	0.7604
$\theta_1^{(1)}$	0.7	0.7237	0.7096	0.7212

Tableau 6: L'estimation de $\phi_1^{(2)}$, $\phi_2^{(2)}$, et $\theta_1^{(1)}$.

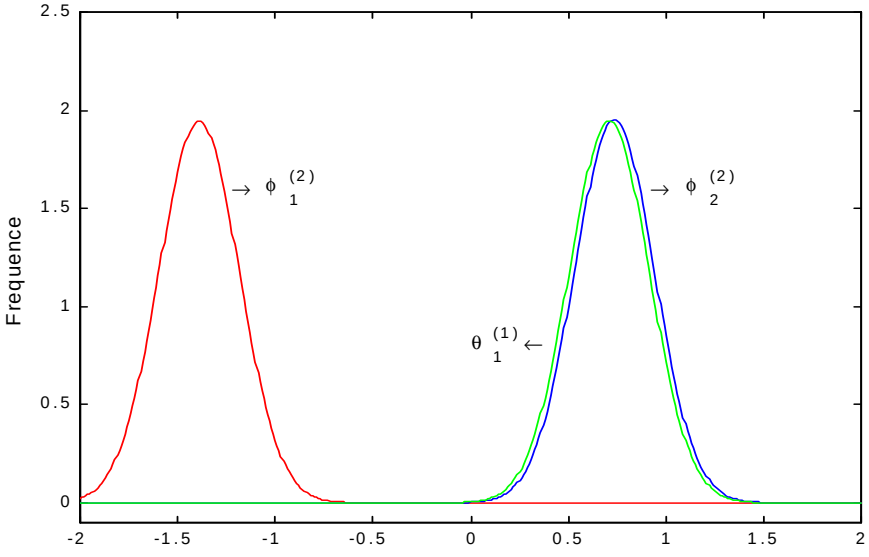


Figure 18: Estimation par noyau des lois MAP des coefficients.

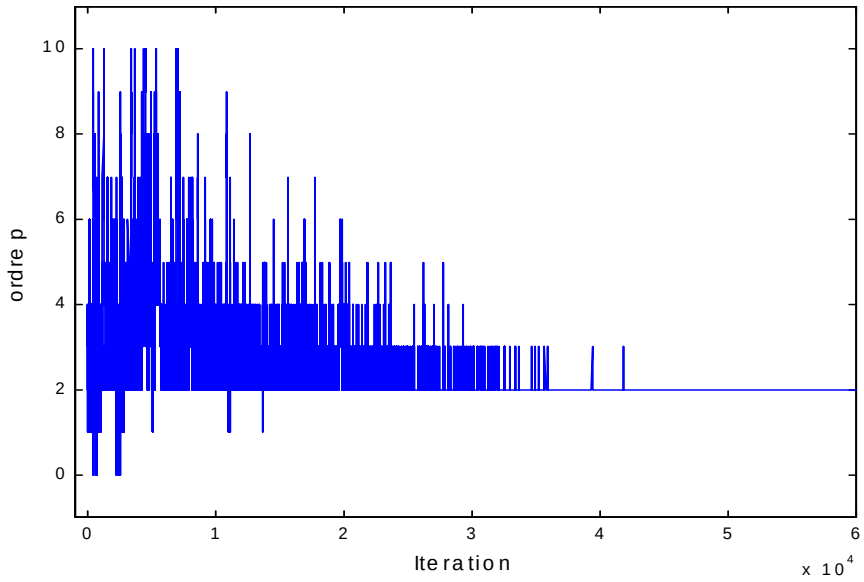


Figure 19a: Evoluation de p au cours de l'Algorithme RS

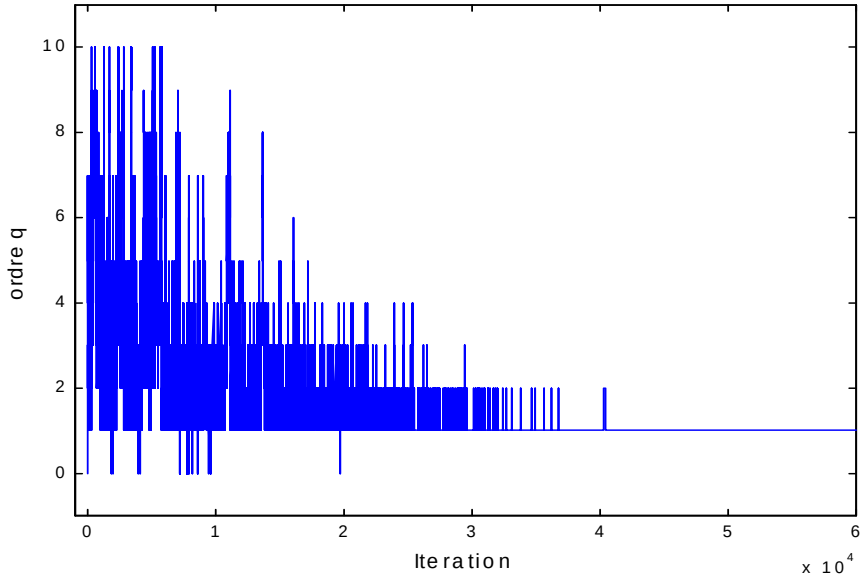


Figure 19b: Evolution de q au cours de l'Algorithme RS

Robustesse de l'estimation

Pour ce dernier exemple, nous étudions la robustesse de nos estimations de p et q. Pour cela, nous avons fait tourner notre algorithme 50 fois. Nous avons calculé la moyenne des histogrammes et l'écart type correspondant. Nous avons représenté dans la figure (20), à gauche, la moyenne moins l'écart type, au milieu la moyenne et à droite, la moyenne plus l'écart type de l'histogramme. Ce dessin montre que notre estimation de p et q est très stable.

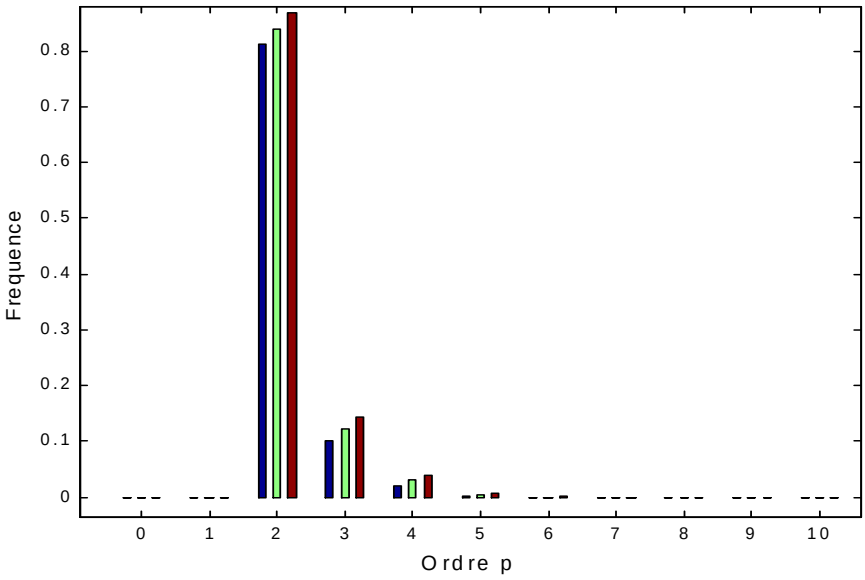


Figure 20a: Robustesse de l'estimation de p

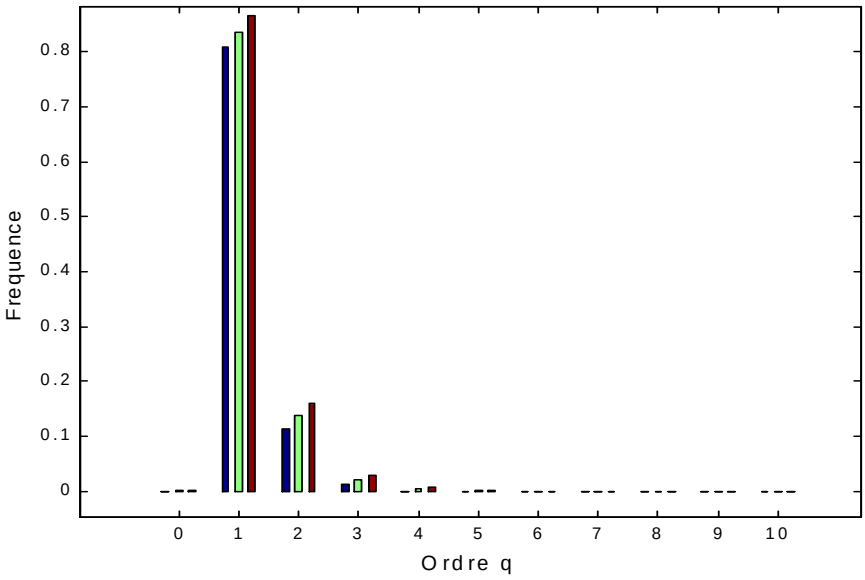


Figure 20b: Robustesse de l'estimation de q

En ce qui concerne l'Algorithme RS, les valeurs obtenues pour p et q sont toujours les bonnes valeurs. Nous avons alors étudié la stabilité des estimations des coefficients du processus ARMA. Ces résultats sont résumés dans le tableau suivant et montrent une très grande stabilité des estimations, autour d'une

valeur proche de la "vrai valeur".

	Valeur vrai	Moyenne Est.	Ecart Est.
$\phi_1^{(2)}$	-1.36	-1.3985	0.005
$\phi_2^{(2)}$	0.7	0.7456	0.0054
$\theta_1^{(1)}$	0.7	0.7254	0.0084

Tableau 7: Stabilité des estimations de $\phi_1^{(2)}$, $\phi_2^{(2)}$, et $\theta_1^{(1)}$.

6 Conclusion

La méthode d'estimation des paramètres d'un ARMA, dont surtout les ordres p et q, par simulation MCMC à Sauts Réversibles donne des résultats satisfaisants, grâce notamment à la reparamètrisation utilisée qui tient compte des conditions de causalité et d'inversibilité. Contrairement aux méthodes classiques, elle permet d'estimer en même temps les deux familles de paramètres: ordre et coefficients.

Il est bien connue (*cf.* [17]) que dans les méthodes bayésiennes, les facteurs de pénalisation qui permettent d'éviter une sur-paramètrisation du modèle, réapparaissent dans les hyperparamètres des lois à priori. Une procédure d'actualisation de ces hyperparamètres nous permet de ne pas figer ce facteur de pénalisation.

L'une des difficultés des méthodes de type MCMC à Sauts Réversibles, est l'absence de critères pour le choix des lois des sauts entre espace de dimension différente. Un travail récent et en cours de publication [9] propose des techniques d'optimisation du choix des paramètres des lois, au sein d'une famille fixée - L'un des prolongements de ce travail serait d'utiliser ces techniques dans notre problème -.

References

[1] H. Akaike, "A new look at the statistical identification model", *IEEE Trans. on Automatic Control*, vol. 19, pp. 716-723, 1974.

[2] M.M. Barbieri and A. O'Hagan, "A reversible jump MCMC sampler for Bayesian analysis of ARMA time series," Technical Report, Università La Sapienza and University of Nottingham. 1996.

[3] O. Barndorff-Nielsen and G. Schou, "On the parametrization of autoregressive models by partial autocorrelation", *J.Multivar. Anal.*, vol. 3, pp. 408-419, 1973.

[4] J.M. Beguin, C. Gourieroux, and A. Monfort, "Identification of an ARMA process: the corner method," dans Time series, Editeur T. Anderson, North-Holland, 1980.

[5] R.J. Bhansali, "The inverse partial correlation function of a time series and its applications," *J.Multivar. Anal.*, vol. 13, pp. 310-327, 1983.

[6] G.E.P. Box and G.M. Jenkins & G.C. Reinsel, *Time Series Analysis: forecasting and control*, Holden Day, 1994.

[7] P.J. Brockwell and R.A. Davis, *Time Series: Theory and Methods*, Springer-Verlag, 1993.

[8] P.J. Brockwell and R.A. Davis, *Introduction to Time series and Forecasting*, Springer-Verlag, 1996.

[9] S.P. Brooks, P Guidici ans G.O. Roberts, *Efficient Construction of Reversible Jump Proposal Distributions*. Technical Report, University of Cambridge. 2001.

[10] S. Geman and D. Geman,"Stochastic Relation, Gibbs Distribution and the Bayesian Restoration of Image," *IEEE Transaction on pattern analysis and machine intelligence*, 6, pp. 721-741, 1984.

[11] W.R. Gilks, S. Richardson and D.J. Spiegelhalter, *Markov Chain Monte Carlo in Practice.* Chapman & Hall *, 1996.*

[12] P.J. Green, "Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination," *Biometrika,* vol. 82, pp. 711-732, 1995.

[13] W.K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, pp. 97-109, 1970.

[14] C.M. Hurvich and C.L. Tsai, "Regression and time series model selection in small samples," *Biometrika*, vol. 76, pp. 297-307, 1989.

[15] M.C. Jones, "Randomly choosing parameters from the stationarity and invertibility region of autoregressive-moving average models," *Appl. Statist.*, vol. 36, pp.134-138, 1987.

[16] S. Kirkpatrick, "Optimization by Simulated Annealing: Quantitative Studies," *Journal of Statistical Physics*, 34, 975-98, 1984.

[17] M. Lavielle, "Optimal Segmentation of Random Processes", *IEEE Trans. on Signal Processing*, Vol. 46, No 46, May 1998.

[18] S.I. Liu, "Comparison of Forecasts for ARMA models between a random coefficient approach and a Bayesian approach," *Comm. Statist.-Meth.*, vol. 24, pp. 319-333, 1995.

[19] J.F. Monahan, "A note on enforcing stationarity in autoregressive-moving average models," *Biometrika*, vol 71, pp. 403-404, 1984.

[20] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, "Equations of state calculations by fast coputing machines," *J. Chemical Physics*, vol. 21, pp. 1087-1091, 1953.

[21] E. Punskaya, C. Andrieu, A. Doucet, and W.J. Fitzgerald, "Bayesian Curve Fitting Using MCMC With Applications to Signal Segmentation," *IEEE Trans. on Sign. Processing*, vol. 50, pp. 747-758, 2000.

[22] M. Quenouille, "Approximate tests of correlation in time series," *Journal of the Royal Statistical Society* B, 11, 68-84, 1949.

[23] C.P. Robert, *The Bayesian Choice. A Decision-Theoretic Motivation*, Springer Texts in Statistics, 1999.

[24] C.P. Robert, *Méthodes de Monte Carlo par Chaîne de Markov*, Economica, 1996.

[25] S. Shaarawy and L. Broemeling, "Inference and prediction with ARMA processes," *Comm. Statist.-Meth.*, vol. 14, pp. 2531-2548, 1985.

[26] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, pp. 461-464, 1978.