Ségmentation bayésienne hiérarchique de processus MA constant par morceaux

S. Suparman

SEGMENTATION BAYESIENNE HIERARCHIQUE DE PROCESSUS MA CONSTANT PAR MORCEAUX

S. Suparman

Université de la Technologie de Yogyakarta
Ringroad Utara Jombor Sleman Yogyakarta Indonésie
e-mail : suparman@netcourrier.com

Résumé

On utilise dans ce travail une méthode bayésienne pour traiter une problème de segmentation de processus MA par morceaux. La complexité des lois a posteriori ainsi que la structure particulière de l'espace des paramètres amène a utiliser la méthode de simulation de type Monte Carlo par Chaînes de Markov à sauts réversibles. Les sorties de l'agorithme sont utilisées pour obtenir plusieurs types d'estimateur des paramètres d'intéret : Maximum Marginal a Posteriori, et Moyenne Marginale a Posteriori.

La délicat problème du réglage des hyperparamètres est contourné en munissant des hyperparamètres de loi, utilisant ainsi une structure bayèsienne hièrarchique.

Mots-Cles : Simulation de Monte Carlo par Chaine de Markov a Sauts Reversibles, Segmentation de processus MA.

Abstract

A bayesian approach is used in this work to deal with study of segmentation of piecewise MA processes. Monte Carlo Markov Chain Simulation methods (with reversible jumps) are used to draw samples distributed according to the posterior distribution of interest. These samples are used to obtain different kinds of parameter estimates : the Marginal Maximum A Posteriori estimate (MAP), the Marginal Minimum Mean Square Error estimate (MMSE) and also the estimated Marginal Posteriors.

Hyperparameters are also considered as random varaiables to face with the difficult problem of choosing hyperparameters.

Keywords : Reversible Jump Markov Chain Monte Carlo Simulation, Segmentation of piecewise MA processes.

1. Introduction

Considérons un signal $y = (y_1, \cdots, y_N)$, ou N est le nombre d'observations, modélisé comme un processus MA avec paramètre constant par morceaux et K $(K_0, \cdots, K_{max})$ ruptures. Ce processus MA est un cas particulier du processus ARMA décrit précédemment lorsque le polynome de la part AR est égal à 1. Alors il est toujours causal. Mathématiquement, le modèle du signal est le suivant :

$$y_n = z_n + \sum_{j=1}^{q_{i,K}} \theta_{i,K,j}^{(q_{i,K})} z_{n-j},\ n \in [[n_{i,K}+1, n_{i+1,K}]],\ i = 0, \cdots, K,$$

où sous l'hypothèse de K rupture : $n_{i,K}$ est le $i^{ème}$ instant de rupture (avec la convention $n_{0,K} = 0$ et $n_{K+1,K} = N$), et our chaque $i^{ème}$ intervalle :

- $q_{i,k}$ et $\theta_{i,K}^{(q_{i,K})} = (\theta_{i,K,1}^{(q_{i,K})}, \cdots, \theta_{i,K,q_{i,K}}^{(q_{i,K})})$ sont l'ordre et le paramètre du processus MA associé à cet intervalle.

- z_n est un bruit gaussien de variance $\sigma^2_{i,K}$ associé au processus MA dans cet intervalle. Soit $z_n \sim N(0,\sigma^2_{i,K})$ pour $n \in [[n_{i,K}+1, n_{i+1,K}]]$.

De plus, on supposera que pour chaque segment i, ce processus MA($q_{i,K}$) verifié la condition d'inversibilité. On sait que (cf. [2]) ce processus MA($q_{i,K}$) est inversible si le vecteur $\theta^{(q_{i,K})}_{i,K}$ appartient à

$$I_{q_{i,K}} = \{\theta^{(q_{i,K})}_{i,K} \in R^{q_{i,K}} \mid \ 1+\theta^{(q_{i,K})}_{i,K,1} z + \cdots + \theta^{(q_{i,K})}_{i,K,q_{i,K}} z^{q_{i,K}} \neq 0,\ z \in C, |z| \leq 1\}$$

Notre but est alors à partir de y, d'estimer le nombre de ruptures k, les instants de rupture $n^{(K)} = (n_{1,K}, \cdots, n_{K,K})$, les ordres $q^{(K)} = (q_{1,K}, \cdots, q_{K,K})$, les paramètres $\theta^{(q_{i,K})}_{i,K}$ et les variance du bruit $(\sigma^2)^{(K)} = (\sigma^2_{0,K}, \cdots, \sigma^{(2)}_{K,K})$.

On utilisera la reparamètrisation avec des fonction d'autocorrélation partielle inverse $\rho^{(q_{i,K})}_{i,K} = (\rho^{(q_{i,K})}_{i,K,1}, \cdots, \rho^{(q_{i,K})}_{i,K,q_{i,K}})$ pour faciliter l'exploration de l'espace de paramètre $\theta^{(q_{i,K})}_{i,K}$ (cf. [1]). Si G est une telle transformation de $\rho^{(q_{i,K})}_{i,K} \in]-1,1[^{q_{i,K}}$ vers $\theta^{(q_{i,K})}_{i,K} \in I_{q_{i,K}}$ et q_{max} est l'ordre maximum du processus MA, en notant $\omega = (K, n^{(K)}, q^{(K)}, \{\theta^{(q_{i,k})}_{i,k}\}^K_{i=0}, (\sigma^2)^{(K)})$ alors la fonction de vraisemblance approchée s'écrit (cf. [9])

$$f(y|\omega) = \prod_{i=0}^{K} (2\pi\sigma^2_{i,K})^{-\frac{1}{2}(n_{i+1,K}-n_{i,K})} \exp - \frac{1}{2\sigma^2_{i,K}} \sum_{n=n_{i,K}+1}^{n_{i+1,K}} (y_n - \sum_{j=1}^{q_{i,K}} G(\rho^{(q_{i,K})}_{i,K}) \hat{z}_{n-j})^2$$

où $\hat{z}_1 = \cdots = \hat{z}_{q_{max}} = 0$ et ou pour n = q_{max}+1, ..., N

$$\hat{z}_n = y_n - \sum_{j=1}^{q_{i,K}} G(\rho^{(q_{i,K})}_{i,K}) \hat{z}_{n-j},\ n \in [[n_{i,K}+1, n_{i+1,K}]]$$

2. Approche Bayésienne

Comme nous adoptons l'approche bayésienne (cf. [8]) pour estimer le paramètre inconnu ω, nous devons donc choisir les distributions a priori sur ω. Notons K_{max} le nombre maximum de rupture et q_{nax} le nombre maximum d'ordre. Les distributions a priori du nombre de rupture et des instants de rupture sont choisis :

- Le nombre de ruptures k suit une distribution binomial B(K_{max}, λ) et
- Les instants de rupture $n^{(K)}$ suivent des lois des postitions d'indice pair de statistique d'ordre à 2K+1 points tirés uniformément sans répétition dans {1, ...,N-1}.

Ensuite pour K et $n^{(K)}$ donnés, à chaque intervalle i (i=0, ..., K) :

- L'ordre $q_{i,K}$ suit une distribution uniforme sur $\{0,...,q_{max}\}$
- Le vecteur coefficient $\rho^{(q_{i,K})}_{i,K}$ suit une distribution uniforme sur l'intervalle $(-1,1)^{q_{i,K}}$
- La variance du bruit $\sigma^2_{i,K}$ suit une distribution gamma inverse $IG(\alpha/2, \beta/2)$.

Nous prenons $\alpha = 2$, λ est choisi uniformément dans (0,1) et la distribution a priori du β suit la loi de Jeffrey.

Notons $\xi = (\lambda, \beta)$. Si $\pi(K|\ \lambda)$ est distribution a priori de K, $\pi(n^{(K)}|\ K)$ la distribution a priori de $n^{(K)}$, $\pi(q^{(K)}|\ K)$ la distribution a priori de $q^{(K)}$, $\pi(\sigma^{2(K)}|\ \alpha, \beta, K)$ est la distribution a priori de $\sigma^{2(K)}$, $\pi(\lambda)$ est la distribution a priori de λ, et $\pi(\beta)$ est la distribution a priori de β alors la distribution a priori du paramètre (ω, ξ) est donnée par :

$$\pi(\omega,\xi) = \pi(K|\ \lambda)\ \pi(q^{(K)}|\ K)\ \pi(q^{(K)}|\ K)\ \pi(\{\rho^{(q_{i,K})}_{i,K}\}^K_{i=0}|\ K, q^{(K)})\ \pi(\sigma^{2(K)}|\ \alpha, \beta, K)\ \pi(\lambda)\ \pi(\beta)$$

Ainsi nous obtenons la distribution a posteriori de (ω, ξ) suivante

$$\pi(\omega,\xi|\ y) = \prod_{i=0}^{K}\ [(2\pi\sigma_{i,K}^2)^{-\frac{1}{2}(n_{i+1,K}-n_{i,K})}\ \exp-\{\frac{1}{2\sigma_{i,K}^2}\ \sum_{n=n_{i,K}+1}^{n_{i+1,K}}\ (y_n - \sum_{j=1}^{q_{i,K}}\ G(\rho_{i,K}^{(q_{i,K})})\hat{z}_{n-j})^2\ ]$$

$$C_{K_{max}}^{K}\lambda^K(1-\lambda)^{K_{max}-K}\ \frac{1}{C_{N-2}^{2K+1}}\prod_{i=0}^{K}\ (n_{i+1}-n_i-1)\beta^{-1}$$

$$\frac{1}{(q_{max}+1)^{K+1}}(\frac{1}{2})^{\sum_{i=0}^{K}q_{i,K}}\ \prod_{i=0}^{K}\ [\frac{1}{\Gamma(\alpha/2)}(\beta/2)^{\beta/2}\ (\sigma_{i,K}^2)^{-\alpha/2-1}\exp-\frac{1}{\sigma_{i,K}^2}(\beta/2)]$$

L'estimation du paramètre ω est basée sur la marginalisation de la loi a posteriori $\pi(\omega,\xi|\ y)$. Ensuite nous utilisons la méthode MCMC (cf. [3], [4], [5] et [6]) pour estimer le paramètre ω.

3. Méthode MCMC

Nous adoptons également un algorithme de Gibbs hybride (cf. [7]) pour simule le choix aléatoire de point selon la loi a posteriori $\pi(\omega,\xi|\ y)$ où $\omega = (K, n^{(K)}, q^{(K)}, \{\theta_{i,k}^{(q_{i,k})}\}_{i=0}^{K}, (\sigma^2)^{(K)}\}$ est le vecteur paramètre et $\xi = (\lambda,\beta)$ le vecteur des hyperparamètre. Cet algoritme se décompose en deux étapes :

- Etape 1 : Simuler selon $\pi(\xi|\ \omega, y)$
- Etape 2 : Simuler selon $\pi(\omega|\ \xi, y)$

La loi $\pi(\xi|\ \omega, y)$ a une forme explicite.

$$\pi(\xi|\ \omega, y) = B(K+1, K_{max}-K+1) \otimes G(\alpha/2(K+1), \sum_{i=0}^{K}\frac{1}{2\sigma_{i,K}^2})$$

Nous pouvons donc employer l'algorithme de Gibbs (cf. [5]) pour la simuler. Par contre, la loi $\pi(\omega|\ \xi, y)$ n'a pas de forme explicite. La simulation exacte étant impossible, nous proposons une étape de l'algorithme hybride qui combine l'algorithme MCMC à sauts reversibles (cf. [6]) et l'algorithme de Gibbs. Cet algorithme se décompose en trois étapes :

- Etape 2.1 : Simuler $\pi(K, n^{(K)}, q^{(K)}, \{\rho_{i,k}^{(q_{i,k})}\}_{i=0}^{K}|\ \xi, y)$
- Etape 2.2 : Simuler $\pi(q_{i,K}, \rho_{i,k}^{(q_{i,k})}|\ K, n^{(K)}, \xi, y)$ pour tout i = 0, ..., K
- Etape 2.3 : Simuler $\pi(\sigma_{i,K}^2|\ K, n^{(K)}, q^{(K)}, \rho_{i,k}^{(q_{i,k})}, \xi, y)$ pour tout i = 0, ..., K

Nous utilisons l'algorithme MCMC à sauts reversibles pour les étapes 2.1 et 2.2, car le nombre du ruptures K et l'ordre $q_{i,K}$ de chaque segment i sont inconnus. D'autre part nous employons l'agorithme de Gibbs pour l'étape 2.3.

4. Résultats de simulation

Un échantillon de taille 250 est simulé. Le nombre de ruptures et les instants de rupture de ce modèle sont respectivement K = 1 et $n^{(1)}$ = 125. L'ordre, les coefficient et la variance du bruit pour chaque segment sont résumés dans le tableau 1.

Tableau 1

$i^{ème}$ segment	$\sigma_{i,1}$	$q_{i,1}$	$\theta_{i,1}^{(q_{i,1})}$
0	0.6	1	0.5610
1	0.4	2	1.4674 0.6256

Le signal origine est représenté figure 1.

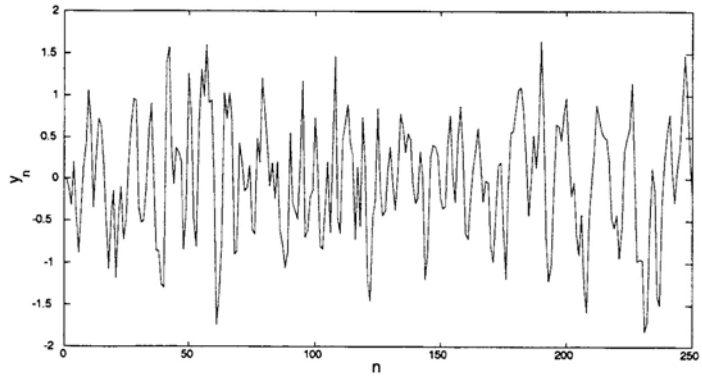

Figure 1

On fait tourner la simulation MCMC pour 60.000 itérations, après une période de chauffage de 10.000 itérations (K_{max} = 10, q_{max}=15). L'histogramme de la distribution marginale a posteriori K est représentée à la figure 2, et nous obtenons le MMAP de K : $\hat{K} = 1$.

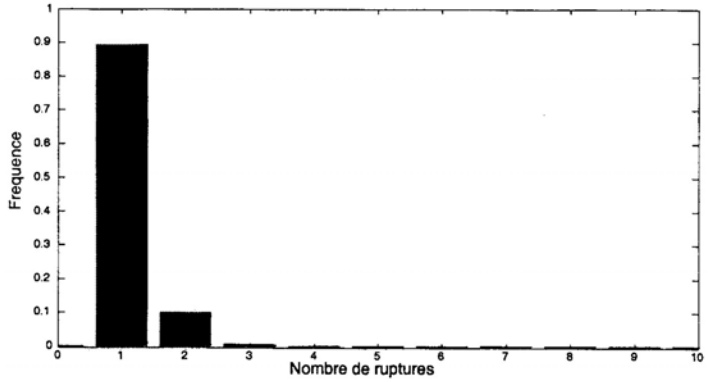

Figure 2

Conditionnellement a $K=\hat{K}$, l'histogramme de la distribution marginale a posteriori de $n^{(K)}$ est donnée dans la figure 3.

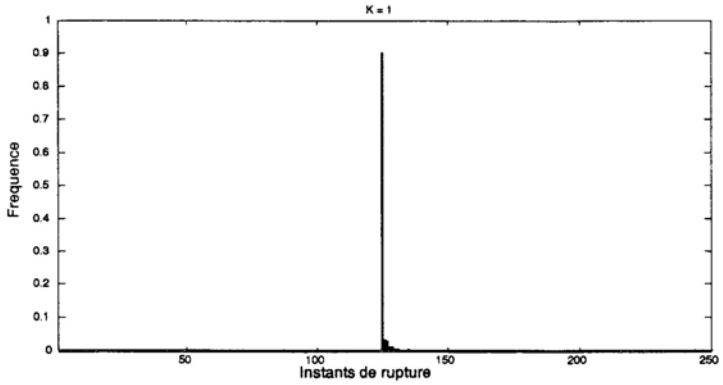

Figure 3

Pour $K=\hat{K}=1$, nous obtenons le Marginal Maximum A posteriori (MMAP) de $n^{(1)}$, $\hat{n}^{(1)}=125$. La superpostion du signal avec le MMAP de $n^{(1)}$ est donnée figure 4

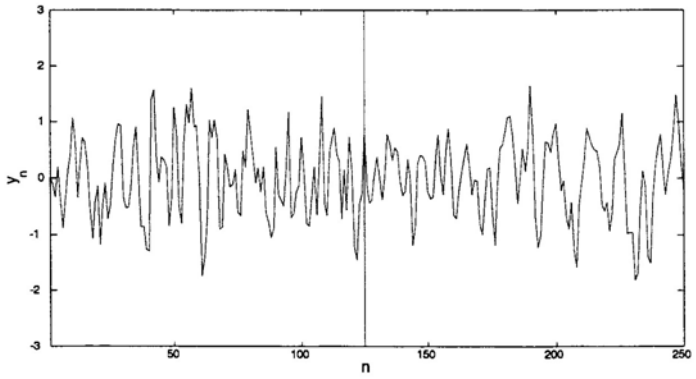

Figure 4

Conditionnellement a K= $\hat{K}$, nous donnons égalemet les histogrammes de la distribution marginale a posteriori de $q_{0,1}$ et $q_{1,1}$ sont données figure 5 et figure 6.

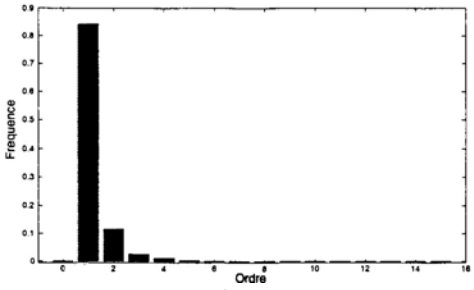

Figure 5

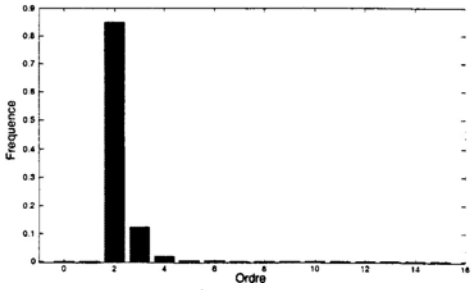

Figure 6

Avec la figure 5 et la figure 6, nous obtenons le MMAP de $q_{0,1}$ et $q_{1,1}$ soit $\hat{q}_{0,1} = 1$ et $\hat{q}_{1,1} = 2$. A $q_{0,1} = \hat{q}_{0,1}$ fixé, la courbe a (figure 7) montre l'estimateur de la densité marginale conditionnelle a posteriori de $\theta_{0,1,1}^{(q_{0,1})}$. De même pour $q_{1,1} = \hat{q}_{1,1}$ fixé, dans la même figure 7 la courbe b et le courbe c montrent l'estimateur de la densité marginale conditionelle a posteriori de $\theta_{1,1,1}^{(q_{1,1})}$ et $\theta_{1,1,2}^{(q_{1,1})}$, utilisant le même noyau gaussien.

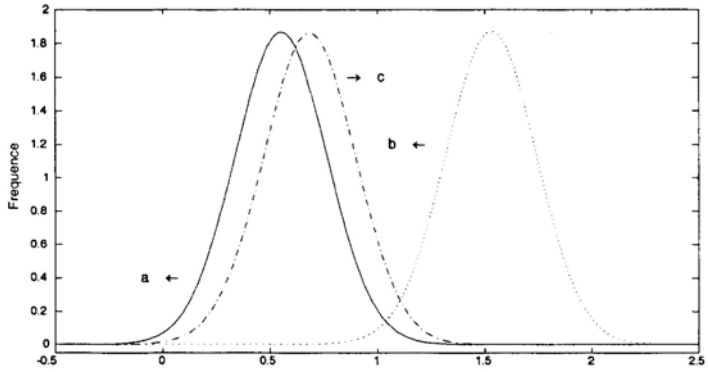

Figure 7

Les histogrammes de la distribution marginal a posteriori de $\sigma_{0,1}$ et $\sigma_{1,1}$ sont donneés figure 8.

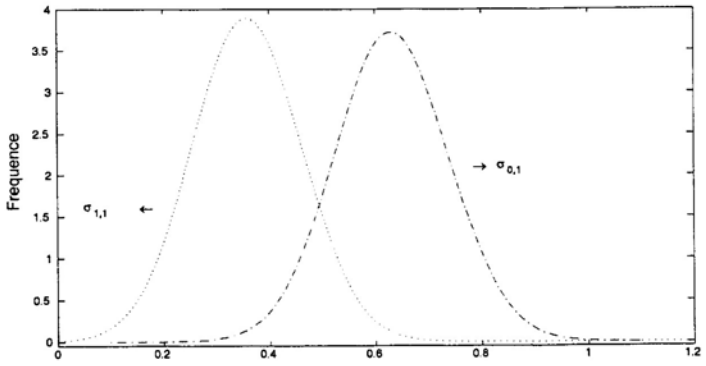

Figure 8

5. Conclusion

Nous avons illustré une applications de la méthode MCMC à sauts réversibles a la segmentation de processus MA par morceaux. L'algorithme donne de bons résultats. Cette technique nous permet d'estimer le nombre de rupture, les instants des ruptures et les paramètres du processus MA. Par contre, la segmentation de processus ARMA par morceaux semble plus délicate par une telle méthode. Peut-être à cause de la lourdeur des calculs.

Bibliographie

[1] Bhansali, R.J. (1983) The inverse correlation function of a time series and its applications, *J.Multivar. Anal*, 13, 310-327.

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

[3] Hastings, W.K. (1970) Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, 57, 97-109.

[4] Metropolis, N., Rosenbluth, A.W., Teller, A.H. and Teller, E. (1953) Equations of state calculations by fast coputing machines, *Journal Chemical Physics*, 21, 1087-1091.

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

[6] Green, P.J. (1995) Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination, *Biometrika*, 82, 711-732.

[7] Robert, C.P. (1996) *Méthodes de Monte Carlo par Chaînes de Markov,* Economica.

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

[9] Shaarawy, S. and Broemeling, L. (1984) Bayesian inferences and forecasts with moving averages processes. *Commun. Statist. – Theory Meth.*, 13(15), 1871-1888.