

Détection bayésienne de saut de fréquence par des méthodes MCMC

M. Vieira, C. Theys et A. Ferrari

I3S, CNRS/UNSA, 41 Bd. Napoléon III, 06041 Nice, France

vieira@unice.fr - theys@unice.fr - ferrari@unice.fr

RÉSUMÉ

Le problème traité est celui de la détection et de l'estimation hors-ligne d'un saut de fréquence dans une sinusoïde complexe bruitée. Une approche bayésienne utilisant une méthode de Monte Carlo par chaînes de Markov (MCMC) nous permet d'estimer l'instant de rupture r et les pulsations avant et après le saut uniquement à partir des observations. Une version adaptative de l'algorithme de Metropolis-Hastings (M-H) à *marche aléatoire* est utilisée permettant de régler de façon optimale et automatique les paramètres de l'algorithme. Les performances de détection estimées en terme de distribution de l'erreur d'estimation de r et d'estimation de l'amplitude de la rupture sont présentées.

ABSTRACT

The problem addressed in this paper is the off-line detection and estimation of frequency jump in a noisy complex sinusoid. A Bayesian approach using Markov chain Monte Carlo methods for estimating the change time r and the frequencies before and after the jump only with observations is proposed. An adaptive version of the random walk Metropolis-Hastings (M-H) algorithm is used allowing to adjust automatically the algorithm parameters. Performances of detection estimated by the distribution of change time estimation error and ones of the estimation of change amplitude are presented.

1 Introduction

Le problème traité est celui de la détection et de l'estimation hors-ligne d'un saut de fréquence dans une sinusoïde complexe bruitée. C'est un problème trouvant de nombreuses applications : systèmes radar et sonar, détection de pannes, ... Si ce problème peut être résolu dans l'espace temps-fréquence, la non linéarité du saut ne permet pas une solution classique basée sur un modèle paramétrique. Une approche bayésienne utilisant les méthodes de Monte Carlo par chaînes de Markov (MCMC) pour estimer l'instant et de l'amplitude de la rupture par moyennage a posteriori est proposée.

Comme la largeur du support des densités a posteriori de ces paramètres dépend du rapport signal sur bruit (RSB) et de l'amplitude du saut, une approche adaptative de l'algorithme de Metropolis-Hastings (M-H) basée sur le principe de régénération de chaîne de Markov présenté par [5] est utilisée.

Les performances de détection et d'estimation du saut de fréquence sont, alors, estimées à partir d'un exemple de simulation.

2 Position du problème

Le signal étudié est une sinusoïde complexe bruitée d'amplitude A , de phase ϕ et de pulsation initiale ω_1 inconnues. La pulsation du signal passe de la valeur ω_1 à la valeur ω_2 inconnue à un instant r inconnu :

$$d_n = \begin{cases} A \exp(j(\omega_1 n + \phi)) + e_n & \text{pour } 0 \leq n \leq r - 1 \\ A \exp(j(\omega_2 n + \phi)) + e_n & \text{pour } r \leq n \leq N - 1 \end{cases} \quad (1)$$

e_n étant un bruit complexe, blanc, gaussien, i.i.d., de moyenne nulle et de variance σ^2 .

La fonction de vraisemblance, étant donné $\mathbf{d} = \{d_0, \dots, d_{N-1}\}$, est donnée par :

$$p(\mathbf{d}|\boldsymbol{\theta}, B, \sigma, I) = \pi^{-N} \sigma^{-2N} \exp\left[-\frac{1}{\sigma^2} \left(\sum_{n=0}^{r-1} |d_n - B \exp(j\omega_1 n)|^2 + \sum_{n=r}^{N-1} |d_n - B \exp(j\omega_2 n)|^2 \right)\right] \quad (2)$$

où $\boldsymbol{\theta} = \{\omega_1, \omega_2, r\}$, $B = A \exp j\phi = B_1 + jB_2$.

Dans le contexte bayésien, le problème est d'obtenir l'expression de la densité de probabilité a posteriori des paramètres inconnus $\boldsymbol{\theta}$ conditionnellement aux données \mathbf{d} et à l'information a priori I . D'après le théorème de Bayes, on a :

$$p(\boldsymbol{\theta}, B, \sigma|\mathbf{d}, I) = \frac{p(\mathbf{d}|\boldsymbol{\theta}, B, \sigma, I)p(\boldsymbol{\theta}, B, \sigma|I)}{p(\mathbf{d}|I)} \quad (3)$$

Comme aucune connaissance a priori sur les paramètres n'est à notre disposition, la densité a priori $p(\boldsymbol{\theta}, B, \sigma|I)$ est non informative : on choisit une densité a priori uniforme pour les paramètres $\boldsymbol{\theta}$ et B telle que $p(\boldsymbol{\theta}, B|I) \propto k$, k étant une constante, et une densité a priori de Jeffreys pour le paramètre σ telle que $p(\sigma|I) \propto 1/\sigma$. La densité a priori des données $p(\mathbf{d}|I)$ est une constante de normalisation (voir [3] pour une discussion approfondie sur l'assignement des densités de probabilité a priori).

La relation (3) peut alors se réécrire :

$$p(\boldsymbol{\theta}, B, \sigma|\mathbf{d}, I) \propto \frac{1}{\sigma} p(\mathbf{d}|\boldsymbol{\theta}, B, \sigma, I) \quad (4)$$

Comme les paramètres B et σ sont, dans notre cas, des paramètres de nuisance, ils sont éliminés par intégration sur leur domaine de définition suivant :

$$p(\boldsymbol{\theta}|\mathbf{d}, I) \propto \int_0^{+\infty} d\sigma \int_{-\infty}^{+\infty} dB_1 \int_{-\infty}^{+\infty} p(\boldsymbol{\theta}, B, \sigma|\mathbf{d}, I) dB_2 \quad (5)$$

Après calcul, la densité de probabilité a posteriori de $\boldsymbol{\theta}$ conditionnellement à \mathbf{d} et I s'écrit :

$$p(\boldsymbol{\theta}|\mathbf{d}, I) \propto \left[\sum_{n=0}^{N-1} |d_n|^2 - \frac{1}{N} \left(\left| \sum_{n=0}^{r-1} d_n \exp(-j\omega_1 n) \right|^2 + \left| \sum_{n=r}^{N-1} d_n \exp(-j\omega_2 n) \right|^2 \right) \right]^{-N+0.5} \quad (6)$$

L'intégration de (6) sur un élément θ_i de l'ensemble $\boldsymbol{\theta}$ est très difficile voire impossible aussi bien de manière analytique que par des techniques d'intégration numériques conventionnelles. C'est pourquoi les algorithmes MCMC, facilement implémentables sont intéressants. En effet, ils permettent d'estimer par échantillonnage toutes les densités conditionnelles des paramètres θ_i $p(\theta_i|\boldsymbol{\theta} - \{\theta_i\}, \mathbf{d}, I)$ sans nécessité la connaissance de la constante de normalisation.

Finalement, l'estimation des paramètres $\hat{\boldsymbol{\theta}}$ est déterminée par la moyenne a posteriori des échantillons générés.

3 Algorithme utilisé

Les algorithmes MCMC ont permis un essor considérable de l'inférence par simulation et en particulier de l'inférence bayésienne. Ils se sont révélés récemment très intéressants sur différents problèmes de traitement du signal comme l'estimation paramétrique, le choix de modèles [7], [8].

Ces méthodes permettent de simuler des densités de probabilité non standard, complexes, multivariées en générant une chaîne de Markov ($x^{(t)}$) ergodique de loi stationnaire la densité d'intérêt invariante $p(\cdot)$ connue à une constante multiplicative près). On a besoin, pour cela, d'un noyau de transition $P(x, dy)$ dont la n^{e} itération converge vers $p(\cdot)$, pour n grand.

3.1 Algorithme de Metropolis-Hastings à marche aléatoire

Une des méthodes MCMC les plus utilisées est celle de Metropolis-Hastings (M-H) [4], pour lequel le noyau de transition est défini par :

$$P_{MH}(x^{(t)}, dy) = q(x^{(t)}, y)\alpha(x^{(t)}, y)dy + \left[1 - \int_R q(x^{(t)}, y)\alpha(x^{(t)}, y)dy \right] \delta_{x^{(t)}}(dy) \quad (7)$$

où $q(x^{(t)}, y)$ est une densité multivariée génératrice de candidat et $\alpha(x^{(t)}, y)$, la probabilité d'acceptation, est définie par :

$$\alpha(x^{(t)}, y) = \begin{cases} \min \left\{ \frac{p(y)q(y, x^{(t)})}{p(x^{(t)})q(x^{(t)}, y)}, 1 \right\} & \text{si } p(x^{(t)})q(x^{(t)}, y) > 0 \\ 1 & \text{ailleurs} \end{cases} \quad (8)$$

On remarque clairement que la probabilité d'acceptation devient indépendante de la densité de probabilité $q(x^{(t)}, y)$ choisie si celle-ci est symétrique.

Pour éviter que $p(x^{(t)}) \neq 0$ i.e. que $\alpha(x^{(t)}, y)$ ne soit pas définie, il suffit que $p(x^{(0)}) > 0$, $x^{(0)}$ étant la valeur initiale de la chaîne de Markov générée.

Une famille importante de densité génératrice de candidat est de la forme $q(x^{(t)}, y) = q_1(y - x^{(t)})$ où $q_1(\cdot)$ est une densité multivariée facilement simulable indépendante de x . Le candidat y est ainsi généré par le processus $y = x^{(t)} + z$ où z est une variable aléatoire d'incrément qui suit $q_1(\cdot)$. Cet algorithme M-H dit à *marche aléatoire* est bien adapté à notre problème puisqu'il ne nécessite pas une connaissance précise du support de la densité cible.

Comme l'on veut simuler 3 densités de probabilité a posteriori conditionnelles $p(\theta_i|\boldsymbol{\theta} - \theta_i, \mathbf{d}, I)$ à partir de la densité conjointe (6), on propose d'utiliser une version *locale* de l'algorithme M-H combinant 3 remises à jour à chaque itération. Dans ce cas, on utilise 3 noyaux de transition : $P_1(\omega_1^{(t)}, dy_1|\omega_2, r)$ produisant y_1 étant donné ω_2 et r , $P_2(\omega_2^{(t)}, dy_2|y_1, r)$ produisant y_2 étant donné y_1 et r et $P_3(r^{(t)}, dy_3|y_1, y_2)$ produisant y_3 étant donné y_1 et y_2 . En effet, il est plus facile de trouver les noyaux de transition convergent vers les différentes densités conditionnelles que de trouver un noyau qui converge vers la densité conjointe. De plus, la convergence est alors plus rapide et plus sûre qu'avec l'algorithme M-H *global*.

Les densités génératrices de candidat $q_i(y_k - \theta_i^{(k)})$ $i = 1, 2, 3$ sont choisies, dans notre cas, monovariées, indépendantes et normales de moyenne nulle et de variance δ_i .

Le choix de δ_i détermine le comportement de la chaîne de Markov sur, au moins, deux points : le taux d'acceptation (i.e. le pourcentage de fois où une nouvelle valeur y_k a été acceptée) et le support couvert par la chaîne. En effet, si δ_i est trop grand, le taux d'acceptation est faible et inversement, si δ_i est trop petit, un très grand nombre d'itérations est nécessaire pour couvrir tout le support de la densité de probabilité à simuler.

Pour l'algorithme M-H *local* à *marche aléatoire* dans le cas unidimensionnel, un taux d'acceptation de 0.5 est recommandé par [1]. On peut, par conséquent, en déduire expérimentalement les valeurs de δ_i c'est à dire les variances optimales des différentes densités génératrices de candidat $q_i(y_k - \theta_i^{(k)})$.

Le tableau ci-dessous donne les valeurs de δ_3 évaluées expérimentalement pour différents RSB et différentes amplitudes de saut de fréquence $\Delta\omega = |\omega_1 - \omega_2|$:

RSB	0 dB	10 dB	20 dB
$ 1 - 1.2 = 0.2$	6	4	2
$ 0.7 - 1.3 = 0.6$	5	3	2
$ 0.9 - 1.9 = 1$	4	2	2

On remarque que la largeur des supports de $p(\theta_i|\boldsymbol{\theta} - \{\theta_i\}, \mathbf{d}, I)$, et en particulier de la densité de probabilité $p(r|\omega_1, \omega_2, \mathbf{d}, I)$, dépend du RSB et de $\Delta\omega$. À chaque réalisation, une étude préliminaire est nécessaire pour déterminer les δ_i optimaux. C'est pourquoi l'utilisation d'une ap-

proche adaptative tenant compte de l'historique de la chaîne de Markov pour construire dynamiquement de meilleures paramétrisations des densités génératrices de candidat s'est révélée intéressante.

3.2 Approche adaptative

La technique d'adaptation utilisée appelée *splitting* et basée sur le concept de régénération de chaîne de Markov est développée dans [5].

Soit $\{X_n : n = 0, 1, \dots\}$, une chaîne de Markov irréductible dans l'espace d'état (E, ε) de noyau de transition $P = P(x, dy)$ et de loi stationnaire invariante $p(\cdot)$.

Si un ensemble $A \in \varepsilon$, tel que $p(A) > 0$ et $P(x, \cdot) = P(y, \cdot)$ pour tous $x, y \in A$ existe, les instants pour lesquels la chaîne entre dans A sont appelés les instants de régénération. A chaque instant de régénération, le noyau de transition est modifié en tenant compte de l'historique de la chaîne de Markov. Il est construit de telle manière que $p(\cdot)$ soit toujours la densité stationnaire invariante cible.

On suppose qu'il est possible de trouver une fonction $s(x)$ et une mesure de probabilité $\nu(dy)$ telles que :

$$p(s) = \int s(x)p(dx) > 0 \quad \text{et} \quad P(x, A) \geq s(x)\nu(A)$$

pour tout $x \in E$ et tout $A \in \varepsilon$, (s, ν) étant appelé *atome* du noyau de transition P .

L'application de cette technique dans le cas de l'algorithme M-H à *marche aléatoire* est décrite dans [6].

Au lieu de chercher un atome pour le noyau de transition (7), [6] propose de trouver (s, ν) tel que :

$$q(x, y)\alpha(x, y)dy \geq s(x)\nu(dy) \quad (9)$$

puisque (7) satisfait la relation

$$P_{MH}(x, dy) \geq q(x, y)\alpha(x, y)dy$$

La régénération de la chaîne de Markov n'est, dans ce cas, possible que lorsque le candidat est accepté. En effet, X_{n+1} et S_n sont alors construits suivant l'algorithme :

1. génération de X_{n+1} à partir de $q(X_n, y)$ avec la probabilité d'acceptation $\alpha(X_n, y)$
2. si rejet du candidat y , alors $S_n = 0$
3. sinon

– génération de S_n à partir de $\mathcal{B}(r(X_n, dy))$ avec

$$r(X_n, dy) = \frac{s(x)\nu(dy)}{q(x, y)\alpha(x, y)}$$

– si $S_n = 1$ alors

– n : instant de régénération

– remise à jour des paramètres de $q(X_n, \cdot)$

3.3 Implémentation

La méthode adaptative de Metropolis-Hastings *locale* à *marche aléatoire* est implémentée suivant l'algorithme ci-dessous :

trouver les valeurs initiales : $\theta^{(0)}$

pour $k = 1 \dots T$ faire :

pour $i = 1, 2, 3$

générer y_k à partir de $q_i(y_k - \theta_i^{(k)})$

générer u à partir de $\mathcal{U}(1, 0)$

$$\alpha_i(\theta_i^{(k)}, y_k) = \min \left\{ 1, \frac{p(y_k | \theta_i^{(k)} - \theta_i^{(k)}, d, I)}{p(\theta_i^{(k)} | \theta_i^{(k)} - \theta_i^{(k)}, d, I)} \right\}$$

si $u \leq \alpha_i(\theta_i^{(k)}, y_k)$ alors

$\theta_i^{(k+1)} = y_k$

calcul de $r(\theta_i^{(k)}, \theta_i^{(k+1)})$

générer S_k à partir de $\mathcal{B}(r(\theta_i^{(k)}, \theta_i^{(k+1)}))$

si $S_k = 1$ alors

remise à jour de δ_i

finsi

sinon

$\theta_i^{(k+1)} = \theta_i^{(k)}$

$S_k = 0$

finsi

fin

fin

estimer $\hat{\theta} = \frac{1}{T-M+1} \sum_{t=M}^T \theta^{(t)}$

À un instant k , la procédure de remise à jour de la variance δ_i de la densité normale génératrice de candidat $q_i(y_k - \theta_i^{(k)})$ tient compte à la fois du taux d'acceptation optimal établi par [1], $a = 0.5$, et de celui entre k et l'instant de régénération précédent, A_i et donc de l'historique de la chaîne de Markov :

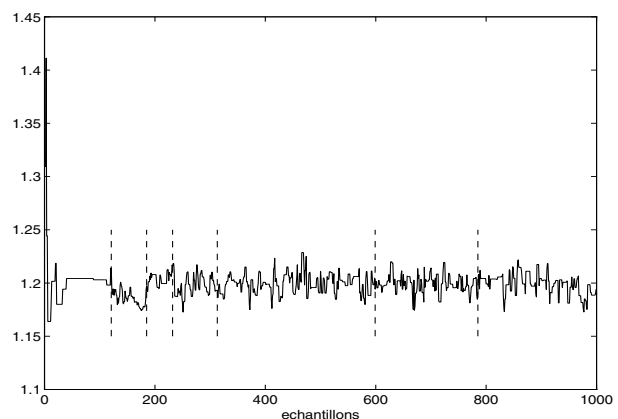
$$\ln(\sqrt{\delta_{i+1}}) = \ln(\sqrt{\delta_i}) + \ln(A_i) - \ln(a) \quad (10)$$

4 Simulations et performances

Des simulations sont effectuées afin d'estimer les performances de la méthode proposée. L'amplitude, la phase et la pulsation avant et après la rupture de la sinusoïde complexe sont respectivement posés à $A = 0.75$, $\phi = \pi/2$, $\omega_1 = 1$ et $\omega_2 = 1.2$. La longueur des signaux générés avec un saut fréquentiel en $r = 50$ est de $N = 100$.

Les variances des densités génératrices de candidat sont initialement posées à $\delta_1 = \delta_2 = 0.5$ et $\delta_3 = 10$.

La figure (1) présentent les 1000 premiers échantillons d'un exemple de chaîne de Markov générée pour estimer ω_2 . Cela permet d'illustrer l'effet de l'adaptation de la variance de la densité génératrice de candidat δ_2 sur la chaîne. Les instants de régénération de δ_2 sont indiqués par les lignes en pointillées.



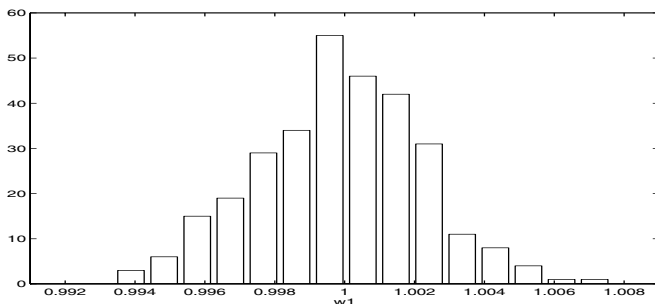
(1) Chaîne de Markov générée pour estimer ω_2

Dans le processus d'échantillonnage, les 2000 premiers échantillons générés par l'algorithme utilisé sont ignorés. Les 3000 suivants permettent alors d'approximer les densités a posteriori $p(\theta_i | \theta - \{\theta_i\}, \mathbf{d}, I)$ puis d'estimer les paramètres. 300 réalisations sont produites pour différents RSB. A chaque réalisation, les paramètres ω_1 , ω_2 et r ont été estimés suivant la méthode proposée.

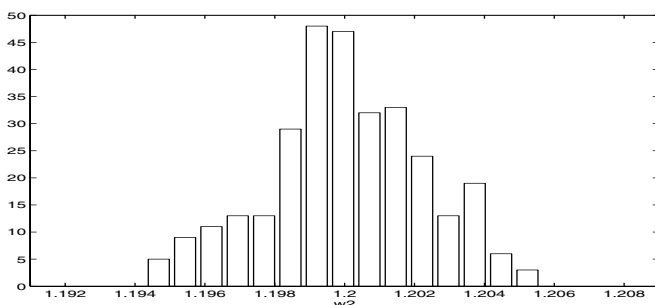
Dans un premier temps, la moyenne et l'écart-type de $\hat{\omega}_1$ et $\hat{\omega}_2$ sont calculés pour les différents RSB afin d'évaluer les performances de l'estimation des pulsations avant et après la rupture. Les résultats obtenus sont présentés dans le tableau ci-dessous :

	RSB (dB)	0 dB	5 dB	10 dB	20 dB
ω_1	Moyenne	0.99868	0.99947	0.99984	0.99983
	Écart-type	9.75e-3	4.49e-3	2.36e-3	8.7e-4
ω_2	Moyenne	1.20224	1.19966	1.20004	1.19997
	Écart-type	1.23e-2	3.85e-3	2.22e-3	8.29e-4

Les figures (2) et (3) présentent les histogrammes de $\hat{\omega}_1$ et $\hat{\omega}_2$ pour un RSB=10dB.

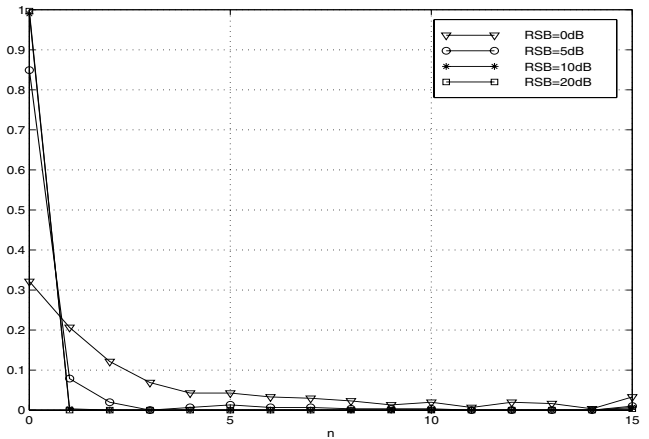


(2) Histogramme de $\hat{\omega}_1$ pour 300 réalisations



(3) Histogramme de $\hat{\omega}_2$ pour 300 réalisations

Les performances de l'estimation hors-ligne de l'instant de rupture ont été calculées en estimant la distribution de l'erreur d'estimation de r c'est à dire $P(\hat{r} = r \pm n)$ avec $n = 0, 1, 2, \dots$ et \hat{r} l'estimation de r (voir [2] et figure (4))



(4) Distribution de l'erreur d'estimation de \hat{r}

5 Conclusion

Une méthode bayésienne de détection et d'estimation hors-ligne d'un saut de fréquence dans une sinusoïde complexe bruitée est proposée utilisant l'algorithme de Metropolis-Hastings local à marche aléatoire adaptatif. Seules les observations sont nécessaires à notre approche. Les performances obtenues sont tout à fait satisfaisantes surtout à partir d'un RSB=10dB où la probabilité d'avoir une estimation correcte de r c'est à dire $P(r = \hat{r})$ est quasiment de 100%.

Références

- [1] W.R. Gilks A. Gelman and G.O. Roberts. *Efficient Metropolis Jumping Rules in Bayesian Statistics 5*, pages 599–608. 1996.
- [2] M. Basseville and I. Nikiforov. *Detection of Abrupt Changes, Theory and Application*. Prentice Hall, 1993.
- [3] G.L. Bretthorst. *Bayesian Spectrum Analysis and Parameter Estimation*. Springer-Verlag, 1989.
- [4] S. Chib and E. Greenberg. Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49 :327–335, 1995.
- [5] E. Nummelin. *General Irreducible Markov Chains and Non-Negative Operators*. Cambridge University Press, 1984.
- [6] L. Tierney P. Mykland and B. Yu. Regeneration in Markov Chain Samplers. *Journal of the American Statistical Association*, 90 :433–441, 1995.
- [7] J.J.K O Ruanaidh and W.J. Fitzgerald. *Numerical Bayesian Methods applied to Signal Processing*. Springer, 1996.
- [8] C. Theys, M. Vieira, and A. Ferrari. Bayesian Estimation of the Parameters of a Polynomial Phase Signal using MCMC methods. *ICASSP*, 5 :3553–3556, 1997.