

# Détection et estimation bayésiennes conjointes de sinusoides noyées dans du bruit

† Christophe Andrieu - Arnaud Doucet - Patrick Duvaut

ENSEA/UCP-ETIS - URA CNRS 2235

6, avenue du Ponceau 95014 Cergy Cedex

andrieu@ensea.fr - douceta@ensea.fr - duvaut@ensea.fr

url : <http://www-rech.ensea.fr/>

## RÉSUMÉ

Dans ce papier nous traitons le problème de détection et estimation conjointes de sinusoides noyées dans du bruit, sous l'angle bayésien. Nous commençons par définir un modèle bayésien original. Ce modèle comporte de nombreux paramètres inconnus, dont le nombre de sinusoides. Tout développement analytique étant impossible, nous proposons un nouvel algorithme stochastique, reposant sur une méthode MCMC (chaînes de Markov Monte Carlo) à sauts réversibles. Nous obtenons de très bons résultats.

## ABSTRACT

In this paper we address the problem of the joint detection and estimation of sinusoids embedded in noise, from a bayesian point of view. We first propose an original bayesian model. A large number of parameters is involved in the model, including the number of sinusoids. No analytical developments can be performed. This lead us to design a new stochastic algorithm relying on reversible jump MCMC (Markov chain Monte Carlo). We obtain very satisfactory results.

## 1 Introduction

Dans de nombreux cas pratiques, le traiteur de signaux est confronté à l'analyse de raies noyées dans du bruit, ce qui en fait l'un des problèmes "canoniques" du domaine. Des approches bayésiennes d'analyse de raies noyées dans du bruit ont été proposées dans [7][10][8]. Dans ces contributions, des approximations analytiques sont souvent effectuées, et le problème de détection/estimation est toujours traité en deux étapes bien distinctes. Dans notre contribution, nous proposons d'une part un modèle bayésien original, et d'autre part un algorithme stochastique, de type Monte Carlo, d'approximation numérique des densités *a posteriori*, traitant ainsi conjointement la détection et l'estimation. Ces densités concentrent la totalité de l'information nécessaire à toute inférence bayésienne.

## 2 Modèle bayésien et objectifs

Ce problème classique peut être formalisé de la manière suivante, en utilisant la représentation en quadrature du signal :

$$\mathbf{y} = \mathbf{D}(\omega_k) \mathbf{a}_k + \mathbf{n}_k \quad (1)$$

où  $\mathbf{D}(\omega_k)$  est la matrice  $N \times 2k$  définie par

$$\mathbf{D}(\omega_k) \triangleq \begin{bmatrix} \mathbf{d}_{c_{1,k}} & \mathbf{d}_{s_{1,k}} & \mathbf{d}_{c_{2,k}} & \dots & \mathbf{d}_{c_{k,k}} & \mathbf{d}_{s_{k,k}} \end{bmatrix}$$

$$\text{où } \mathbf{d}_{c_{j,k}} \triangleq \begin{pmatrix} 1 & \cos[\omega_{j,k}] & \dots & \cos[\omega_{j,k}(N-1)] \end{pmatrix}^t$$

$$\mathbf{d}_{s_{j,k}} \triangleq \begin{pmatrix} 0 & \sin[\omega_{j,k}] & \dots & \sin[\omega_{j,k}(N-1)] \end{pmatrix}^t$$

$$\mathbf{a}_k \triangleq \begin{pmatrix} a_{1,k}^c & a_{1,k}^s & \dots & a_{k,k}^c & a_{k,k}^s \end{pmatrix}^t$$

$$\omega_k \triangleq \left( \omega_{1,k} \quad \omega_{2,k} \quad \dots \quad \omega_{k-1,k} \quad \omega_{k,k} \right)^t \quad (3)$$

$\mathbf{a}_k$  et  $\omega_k$  contiennent respectivement les amplitudes des cosinus et sinus, les pulsations des  $k$  sinusoides.  $\mathbf{y}$  est un vecteur contenant  $N$  observations et  $\mathbf{n}_k$  est un bruit additif blanc gaussien centré, de variance  $\sigma_k^2$ . On peut toutefois lever ces hypothèses [1]. Nous introduisons la notation  $\theta_k = (\mathbf{a}_k^t, \omega_k^t, \sigma_k^2)^t$ . Ne disposant que des observations  $\mathbf{y}$ , l'objectif est d'estimer les inconnues  $k$  et  $\theta_k$ .

Dans un cadre bayésien, il est nécessaire d'assigner à chacun des paramètres et hyperparamètres introduits une densité de probabilité *a priori*. On peut montrer que sous les contraintes  $\Lambda \triangleq \mathbb{E}[k]$  et  $\Xi \triangleq \frac{1}{N} \mathbb{E}[\mathbf{a}_k^t \mathbf{D}^t(\omega_k) \mathbf{D}(\omega_k) \mathbf{a}_k]$ , un argument du type maximum d'entropie mène naturellement à la densité *a priori* conjointe suivante [3] :

$$p(k, \mathbf{a}_k, \omega_k) = \frac{\Lambda^k}{k!} \exp(-\Lambda) \frac{|\mathbf{D}^t(\omega_k) \mathbf{D}(\omega_k)|^{1/2}}{(2\pi\delta^2)^{k/2}} \times$$

$$\times \exp\left[ \frac{\mathbf{a}_k^t \mathbf{D}^t(\omega_k) \mathbf{D}(\omega_k) \mathbf{a}_k}{-2\delta^2} \right] \frac{\mathbb{I}_{\Omega_k}(\omega_k)}{\pi^k}$$

( $\Omega_k \triangleq \{\omega_k \in [0, \pi]^k; 0 \leq \omega_{1,k} < \omega_{2,k} < \dots < \omega_{k,k} < \pi\}$ ) est un ensemble sur lequel ne se posent pas de problème d'identifiabilité du modèle, et  $\delta^2 \triangleq N\Xi/\Lambda$ . Nous verrons lors de la phase expérimentale comment ces hyperparamètres peuvent être réglés de manière objective à partir des données. Le paramètre  $\sigma_k^2$ , qui est un paramètre d'échelle, se voit assigné l'*a priori* non informatif impropre classique, dit de Jeffreys [6],  $p(\sigma_k^2) \propto 1/\sigma_k^2$ , et est supposé indépendant des autres paramètres, dont la dimension  $k$ .

Toute inférence bayésienne est alors basée sur la connaissance de la densité *a posteriori* suivante :

$$p(\theta_k, k | \mathbf{y}) \propto p(\mathbf{y} | \theta_k, k) p(\theta_k, k) \quad (4)$$

†Ordre alphabétique

dont il est impossible d'obtenir une expression analytique : (a) de la constante de normalisation, (b) de toute densité marginale, (c) du maximum *a posteriori*. Cela motive les deux sections suivantes, c'est-à-dire la mise au point d'un algorithme stochastique efficace d'estimation de cette densité.

### 3 MCMC : principe

Les méthodes MCMC consistent à générer une chaîne de Markov  $(\mathbf{x}_k)_{k \in \mathbb{N}}$ , dont la densité d'équilibre  $\pi(\cdot)$  est la densité dont on cherche à connaître certaines caractéristiques. Lorsque cette chaîne de Markov a convergé vers sa densité d'équilibre, il est licite (sous certaines conditions généralement vérifiées) d'écrire le type de relation (loi des grands nombres) (pour  $f$   $\pi$ -intégrable) :

$$\lim_{n \rightarrow +\infty} \frac{1}{n+1} \sum_{k=0}^n f(\mathbf{x}_k) = \mathbb{E}_\pi(f(\mathbf{x}))$$

Une introduction simple des méthodes de construction de ce type de chaînes de Markov, en français et comportant des références clefs, est donnée dans [1]. [12] est une référence bien plus exhaustive, mais pouvant se révéler ardue au premier abord. Rappelons toutefois que, grossièrement, les MCMC sont la succession de propositions aléatoires de nouvelles valeurs pour l'ensemble ou une partie des paramètres (au moyen d'une densité dite instrumentale), suivies d'étapes d'acceptations/réjections *ad hoc*.

### 4 Algorithme

Il est impossible d'expliquer en détail l'ensemble des étapes de notre algorithme en moins de quatre pages. C'est pourquoi nous n'en donnons que les grandes lignes et idées principales, et invitons le lecteur intéressé à se référer à [3] pour plus de précisions.

L'algorithme effectue en boucle l'ensemble des étapes suivantes :

1. Initialisation aléatoire des paramètres.
2. Mise à jour successive, pour  $k$  fixé, des paramètres  $\mathbf{a}_k$ ,  $\omega_k$  et  $\sigma_k^2$ .
3. Choisir aléatoirement un des "mouvements" proposant un changement de dimension :
  - (a) "naissance" d'une sinusoïde (dont les paramètres sont proposés aléatoirement).
  - (b) "mort" d'une sinusoïde (victime choisie au hasard).
  - (c) "scission" aléatoire d'une sinusoïde en deux sinusoïdes proches.
  - (d) "fusion" aléatoire de deux sinusoïdes adjacentes en une seule.
4. Aller en 2.

La mise à jour des paramètres  $\mathbf{a}_k$  et  $\sigma_k^2$  est une chose très classique sur laquelle nous n'insistons pas. En revanche la mise à jour de  $\omega_k$  est beaucoup moins évidente, du fait des fortes non-linéarités introduites par ce paramètre. Nous avons donc choisi une mise à jour "d'un paramètre à la fois", c'est-à-dire que chacun des  $\omega_{i,k}$  est mis à jour conditionnellement

aux autres valeurs  $\omega_k \setminus \{\omega_{i,k}\}$  (ainsi que  $\mathbf{a}_k$  et  $\sigma_k^2$ ). La densité de proposition est le mélange d'une loi élaborée à partir du périodogramme de l'observation à laquelle on a retiré les composantes du signal correspondant à  $\omega_k \setminus \{\omega_{i,k}\}$  et d'une marche aléatoire autour de la valeur courante.

Une propriété essentielle des chaînes de Markov, la réversibilité, permet d'assurer simplement que la densité d'équilibre de la chaîne de Markov est bien la densité désirée. Ceci est crucial, et n'est pas une chose évidente lorsque la taille des paramètres n'est pas fixée. C'est ce qui a motivé [11], dont nous appliquons ici les résultats permettant d'imposer cette propriété. Les mouvements de "mort" et "naissance" se conçoivent relativement bien [3]. En revanche, les mouvements "scission" et "fusion" demandent plus d'attention. Ce type de mouvement, destiné à lever l'ambiguïté entre les configurations  $C_1$  "une grande sinusoïde"  $(a^c, a^s, \omega)$  et  $C_2$  "deux petites sinusoïdes très proches"  $(a_i^c, a_i^s, \omega_i)$  pour  $i = 1, 2$ , doit pallier la difficulté qu'il y a à passer de  $C_1$  à  $C_2$  au moyen d'un décès suivi de deux naissances. Le principe en est très simple :

Proposition de passage de  $C_1$  à  $C_2$  :

1. Choisir une sinusoïde au hasard : configuration  $C_1$ .
2. Répartir aléatoirement l'énergie  $E$  de  $C_1$  en deux quantités  $E_1$  et  $E_2$ .
3. Subdiviser aléatoirement  $E_i$  en deux de manière à obtenir  $(|a_i^c|, |a_i^s|)$  pour  $i = 1, 2$ .
4. Tirer aléatoirement les 4 signes de  $(a_i^c, a_i^s)$  pour  $i = 1, 2$ .
5. Tirer aléatoirement  $\omega_i$  ( $i = 1, 2$ ) de manière à ce que  $\omega$  soit barycentre de  $(\omega_i, E_i)$  pour  $i = 1, 2$ .

Proposition de passage de  $C_2$  à  $C_1$  :

1. Choisir deux sinusoïdes adjacentes au hasard : configuration  $C_2$ .
2. Calculer  $E \triangleq (a_1^c)^2 + (a_1^s)^2 + (a_2^c)^2 + (a_2^s)^2 \triangleq E_1 + E_2$ .
3. Subdiviser aléatoirement  $E$  en deux, de manière à obtenir  $(|a^c|, |a^s|)$ .
4. Tirer aléatoirement les 2 signes de  $(a^c, a^s)$ .
5. Calculer  $w$  barycentre de  $(\omega_i, E_i)$  pour  $i = 1, 2$ .

On constate, qu'étant donnés tous les paramètres  $(a^c, a^s, \omega)$   $(a_i^c, a_i^s, \omega_i)$  pour  $i = 1, 2$  (cela inclut implicitement les variables aléatoires ayant permis les scissions d'énergie et les divers choix de signes, selon bien sûr la configuration de départ,  $C_1$  ou  $C_2$ ), on a défini une transformation déterministe inversible entre  $C_1$  et  $C_2$ . C'est le principe mis en lumière par [11], et qui permet d'assurer la propriété de réversibilité lorsque la dimension des paramètres n'est pas fixée.

La convergence de l'algorithme peut être établie relativement simplement, nous renvoyons le lecteur à [3]

### 5 Résultats expérimentaux

En pratique, et comme nous l'avons souligné dès la section 2, il est nécessaire de fixer les valeurs de  $\Lambda$  et  $\Xi$ . En fait, on constate en pratique que seul l'ordre de grandeur de  $\delta^2$  va être nécessaire, et nous verrons pourquoi. Une solution consiste donc à considérer dans un premier temps  $\delta^2$  et  $\Lambda$  comme des hyperparamètres à estimer au même titre que les autres. Ainsi

le modèle bayésien se trouve modifié, et on estime finalement la densité *a posteriori* suivante :

$$\begin{aligned} p(\delta^2, \Lambda, \theta_k, k | \mathbf{y}) &\propto p(\mathbf{y} | \delta^2, \Lambda, \theta_k, k) p(\delta^2, \Lambda, \theta_k, k) \\ &\propto p(\mathbf{y} | \delta^2, \Lambda, \theta_k, k) \times \\ &\quad \times p(\theta_k, k) p(\delta^2) p(\Lambda) \end{aligned} \quad (5)$$

On assigne au paramètre d'échelle  $\delta^2$  l'*a priori* diffus

$$\delta^2 \sim \mathcal{F}\mathcal{G}(\alpha, \beta) \quad (\alpha = 2, \beta = \mathbf{y}^s \mathbf{y}) \quad (6)$$

$\delta^2$  peut être aisément remis à jour, puisqu'il s'agit d'un *a priori* conjugué [12]. Pour le paramètre  $\Lambda$  on applique la même méthode, mais ici on impose un *a priori* quasi non informatif [6] :

$$\Lambda \sim \mathcal{F}\mathcal{G}(1/2 + \varepsilon_1, \varepsilon_2) \quad (\varepsilon_i \ll 1 \quad i = 1, 2) \quad (7)$$

Dans les faits on obtient rapidement une estimation de  $p(\delta^2 | \mathbf{y})$  et  $p(\Lambda | \mathbf{y})$ , fournissant des ordres de grandeurs pour  $\delta^2$  et  $\Lambda$ . Il suffit alors d'appliquer l'algorithme pour des valeurs fixées  $\delta_0^2$  et  $\Lambda_0$  (la moyenne par exemple), pour estimer (4). En pratique, les estimations des paramètres  $\mathbf{a}_k$  et  $\omega_k$  sont meilleures que dans le cas où l'on estime (5).

Nous présentons deux types d'expériences semblables à [8]. La première a pour paramètres :  $N = 64, k = 3, E_i \triangleq a_i^{c2} + a_i^{s2}$

$i$	$E_i$	$-\arctan(a_i^s/a_i^c)$	$\omega_i/2\pi$
1	20	0	0.2
2	6.3246	$\pi/4$	$0.2 + 1/N$
3	20	$\pi/3$	$0.2 + 2/N$

Avec comme définition pour le rapport signal/bruit  $RSB = 10 \log_{10} E_1 / (2\sigma^2)$ , nous avons appliqué notre algorithme pour trois expériences pour lesquelles  $RSB_a = 3db, RSB_b = 0db$  et  $RSB_c = -3db$ . Nous présentons les densités *a posteriori* des différents paramètres dans chacun des cas. Nous ne présentons les estimations de  $p(\delta^2 | \mathbf{y})$  et  $p(\Lambda | \mathbf{y})$  que pour  $RSB_b$ . On note la dégradation des performances lorsque le  $RSB$  diminue : de l'énergie peut être attribuée à tort à une sinusoïde. Le tableau 1 donne les quantités permettant de calculer les facteurs de Bayes [6].

La deuxième expérience a pour paramètres :  $N = 64, k = 2$

$i$	$E_i$	$-\arctan(a_i^s/a_i^c)$	$\omega_i/2\pi$
1	20	0	0.2
2	20	$\pi/4$	$0.2 + 1/(lN)$

Avec la même définition que précédemment,  $RSB_l = 3db$  est fixé, et  $l = 1, 4, 6, 8$ . Le tableau 2 donne les quantités permettant de calculer les facteurs de Bayes correspondant.

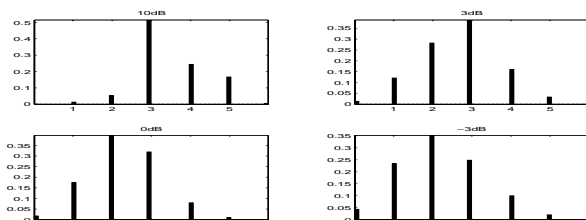


FIG. 1 —  $\hat{p}(k | \mathbf{y})$  pour 4  $RSB$

On note la richesse des résultats et la finesse de l'information qu'ils fournissent sur les données. Cependant, afin d'évaluer les performances statistiques du modèle bayésien et de

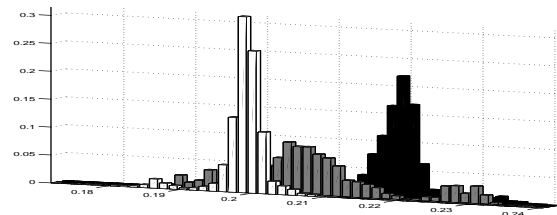


FIG. 2 —  $\hat{p}(\omega_i/2\pi | \mathbf{y}, k = 3)$  pour  $i = 1, 2, 3$  et  $RSB_a$

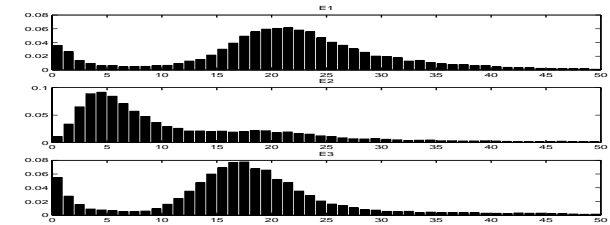


FIG. 3 —  $\hat{p}(E_i | \mathbf{y}, k = 3)$  pour  $i = 1, 2, 3$  et  $RSB_a$

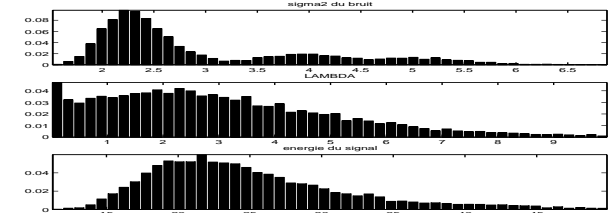


FIG. 4 —  $\hat{p}(\delta^2 | \mathbf{y}), \hat{p}(\Lambda | \mathbf{y})$  et  $\hat{p}(\sigma^2 | \mathbf{y})$  pour  $RSB_b$

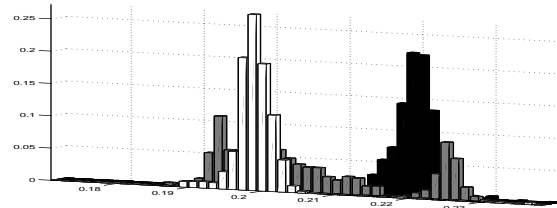


FIG. 5 —  $\hat{p}(\omega_i/2\pi | \mathbf{y}, k = 3)$  pour  $i = 1, 2, 3$  et  $RSB_b$

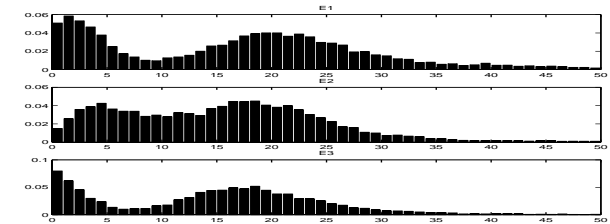


FIG. 6 —  $\hat{p}(E_i | \mathbf{y}, k = 3)$  pour  $i = 1, 2, 3$  et  $RSB_b$

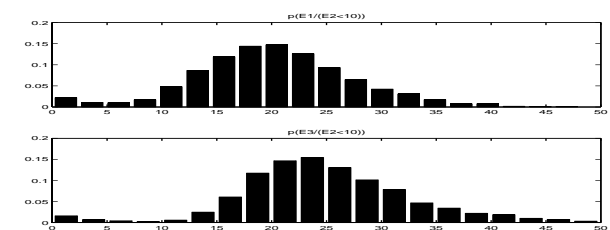


FIG. 7 —  $\hat{p}(E_i | E_2 < 10, k = 3)$  pour  $i = 1, 3$  et  $RSB_b$

$RSB/k$	0	1	2	3	4	5	6
10dB	0.0020	0.0045	0.0121	0.1144	0.0723	0.0825	0.0042
3dB	0.0131	0.0392	0.0589	0.0785	0.0418	0.0142	0.0013
0dB	0.0179	0.0698	0.1258	0.1210	0.0481	0.0120	0.0007
-3dB	0.0439	0.1050	0.1414	0.1345	0.0971	0.0448	0.0076

TAB. 1 —  $\hat{p}(k | \mathbf{y}) / p(k)$  pour 4  $RSB$

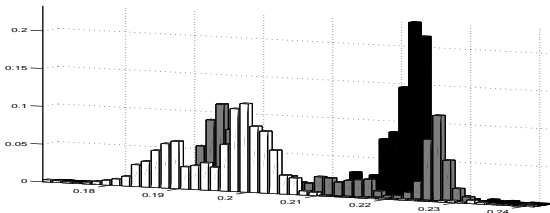


FIG. 8 —  $\hat{p}(\omega_i/2\pi | \mathbf{y}, k = 3)$  pour  $i = 1, 2, 3$  et  $RSB_c$

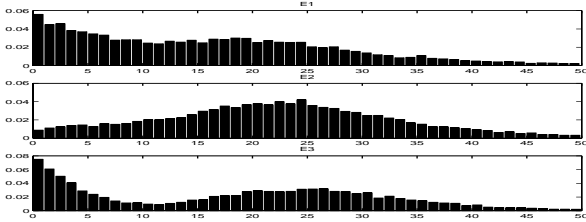


FIG. 9 —  $\hat{p}(E_i | \mathbf{y}, k = 3)$  pour  $i = 1, 2, 3$  et  $RSB_c$

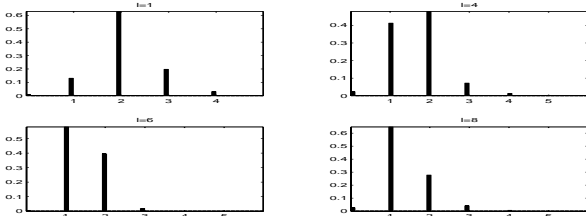


FIG. 10 —  $\hat{p}(k | \mathbf{y})$  pour  $l = 1, 4, 6, 8$  et  $RSB_d$

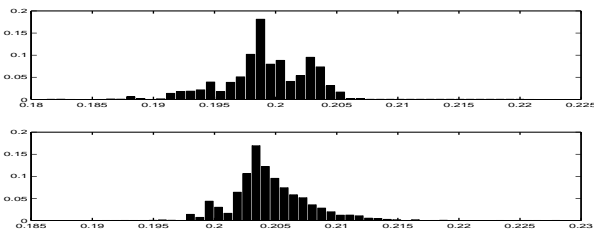


FIG. 11 —  $\hat{p}(\omega_i/2\pi | \mathbf{y}, k = 2)$  pour  $l = 4, i = 1, 2$  et  $RSB_d$

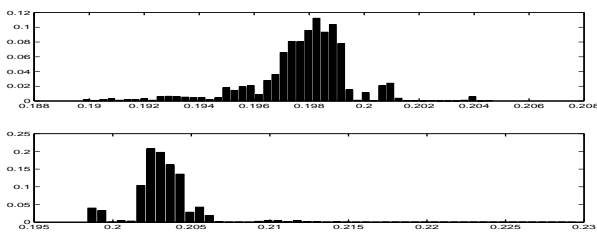


FIG. 12 —  $\hat{p}(\omega_i/2\pi | \mathbf{y}, k = 2)$  pour  $l = 6, i = 1, 2$  et  $RSB_d$

$\Gamma \setminus k$	0	1	2	3	4	5	6
1	0.0108	0.0547	0.2184	0.0859	0.0223	0.0002	0.0000
4	0.0227	0.2411	0.3262	0.0861	0.0372	0.0003	0.0011
6	0.0056	0.3612	0.3059	0.0275	0.0001	0.0004	0.0017
8	0.0257	0.405	0.2174	0.0602	0.0252	0.0005	0.0001

TAB. 2 —  $\hat{p}(k | \mathbf{y}) / p(k)$  pour  $l = 1, 4, 6, 8$

l’algorithme associé, il est nécessaire d’effectuer une étude (Monte Carlo !) sur de nombreuses réalisations de l’observation. Nous invitons le lecteur intéressé à se référer à [3] pour cette étude.

## 6 Conclusion

Dans cette contribution, nous proposons une solution bayésienne au problème de l’analyse spectrale de raies noyées dans du bruit. Nous élaborons un modèle bayésien original ainsi qu’un algorithme stochastique associé, reposant sur les MCMC à sauts réversibles. Cet algorithme permet de traiter de manière conjointe le problème de détection et estimation sur un modèle qui n’autorise pas de développements analytiques exacts. Les résultats obtenus sont remarquables, tant du point de vue de la détection que de l’estimation des divers paramètres, et cela dans des situations difficiles, pour lesquelles les raies sont proches et les observations très bruitées. De plus, nous montrons comment les divers hyperparamètres introduits dans la modélisation, peuvent être aisément estimés à partir des données, limitant au maximum tout arbitraire.

## Références

- [1] † C. Andrieu, A. Doucet et P. Duvaut, “Méthodes Monte Carlo par chaînes de Markov appliquées au traitement du signal”, Rap. tech. ENSEA/UCP-ETIS #2, mars 1997. (<http://www-rech.ensea.fr/>).
- [2] † C. Andrieu, A. Doucet and P. Duvaut, “Bayesian Estimation of Filtered Point-processes using MCMC”, in Proc. Conf. IEEE Asilomar’97, November 1997.
- [3] † C. Andrieu, A. Doucet and P. Duvaut, “Joint Bayesian Model Selection and Estimation of Noisy Sinusoids via Reversible Jumps MCMC”, submitted 97.
- [4] † C. Andrieu, A. Doucet and P. Duvaut, “Robust Bayesian Harmonic Retrieval using MCMC”, submitted, 1997.
- [5] C. Andrieu and P. Duvaut, “Bayesian Detection and Estimation of Shot-noise Processes using Reversible Jump”, in Proc. Conf. IEEE ICASSP’ 97, Munich, March 1997.
- [6] J.O. Berger, Statistical Decision Theory and Bayesian Analysis, Springer-Verlag 1985.
- [7] G.L. Bretthorst, Bayesian Spectrum Analysis and Parameter Estimation, Springer-Verlag, New-York, 1988.
- [8] P.M. Djurić, “A Model Selection Rule for Sinusoids in White Gaussian Noise”, IEEE Trans. Sig. Proc., vol 44, no. 7, pp. 1744-1751.
- [9] A. Doucet and P. Duvaut, “Bayesian Estimation of Instantaneous Frequency”, in Proc. Work. IEEE-TFTS’96, Paris, June 1996.
- [10] L. Dou and R.J.W. Hodgson, “Bayesian inference and Gibbs sampling in spectral analysis and parameter estimation : I & II”, Inverse Problems, 1995 and 1996.
- [11] P.J. Green, “Reversible Jump MCMC Computation and Bayesian Model Determination”, Biometrika, vol. 82, pp. 711-732, 1995.
- [12] C.P. Robert, Méthodes de Monte Carlo par chaînes de Markov, Economica, 1996.