

Segmentation conjointe de données poissonniennes

Nicolas DOBIGEON, Jean-Yves TOURNERET

IRIT/ENSEEIH/TSASA

2 rue Charles Camichel, BP 7122, 31071 Toulouse cedex 7, France

Nicolas.Dobigeon@enseeiht.fr, Jean-Yves.Tourneret@tesa.prd.fr

Résumé – Cet article étudie une méthode de segmentation conjointe de plusieurs signaux issus d’une application astrophysique. Les signaux considérés possèdent des lois de Poisson dont les paramètres sont constants sur des intervalles à déterminer. La méthode proposée est basée sur une analyse Bayésienne hiérarchique. Une méthode de simulation basée sur l’échantillonneur de Gibbs permet d’estimer les paramètres et hyperparamètres inconnus intervenant dans le modèle de segmentation conjointe. Des simulations effectuées sur des données synthétiques illustrent les performances de cette méthode de segmentation conjointe.

Abstract – This paper addresses the problem of detecting significant intensity variations in multiple astronomical time-series. This detection is achieved by using a constant Poisson rate model and a hierarchical Bayesian approach. An appropriate Gibbs sampling strategy allows joint estimation of the unknown parameters and hyperparameters. Simulation results performed on synthetic data illustrate the performance of the proposed algorithm.

1 Introduction

Segmenter un signal ou une image est un problème qui a reçu un vif intérêt dans la communauté du traitement du signal et des images (voir par exemple [1] et [2]). De nombreuses méthodes basées sur des approches Bayésiennes ou sur le principe du maximum de vraisemblance pénalisé permettent de résoudre ces problèmes de segmentation [3, 4, 5]. Nous nous intéressons dans cet article à un problème de segmentation de données astrophysiques présenté initialement dans [6]. La méthode récursive proposée dans [6] consistait à décomposer le signal d’intérêt en deux blocs par optimisation d’un critère approprié et de recommencer l’opération sur chacun des blocs ainsi créés jusqu’à ce qu’un critère d’arrêt soit vérifié. L’avantage de la méthode proposée est sa simplicité due au fait qu’on cherche à chaque itération une seule rupture. Par contre, le principal inconvénient de cette approche est la nécessité d’utiliser une règle d’arrêt qui influence fortement les résultats de segmentation. Cet article propose une nouvelle méthode de segmentation basée sur une approche Bayésienne hiérarchique. Cette méthode ne nécessite pas de définir une règle d’arrêt et permet de segmenter conjointement plusieurs signaux issus de capteurs différents. On suppose comme dans [6] que les propriétés statistiques des signaux observés sont définies par des lois de Poisson de paramètres constants par morceaux :

$$y_{j,i} \sim \mathcal{P}(\lambda_{j,k}),$$

avec $j = 1, \dots, J$, J désignant le nombre de signaux à segmenter, $i \in I_{j,k} = [l_{j,k-1} + 1, l_{j,k}[$, $k = 1, \dots, K_j$, où $I_{j,k}$ est le $k^{\text{ième}}$ intervalle du $j^{\text{ième}}$ signal, $l_{j,k}$ est le point après lequel a lieu la $k^{\text{ième}}$ rupture dans le $j^{\text{ième}}$ signal (avec $l_{j,0} = 0$ et $l_{j,K_j} = n$, où n est le nombre de points de chaque séquence observée) et K_j est le nombre de ruptures dans le $j^{\text{ième}}$ signal. Enfin la notation $\mathcal{P}(\lambda_{j,k})$ désigne une loi de Poisson de paramètre $\lambda_{j,k}$. On suppose de plus que les séquences $y_k = (y_{k,1}, \dots, y_{k,n})$ et $y_l = (y_{l,1}, \dots, y_{l,n})$ sont indépendantes pour $k \neq l$. Segmenter conjointement les signaux $y_j = (y_{j,1}, \dots, y_{j,n})$ consiste à esti-

mer les nombres de ruptures K_j et les positions associées $l_{j,k}$ (pour $j = 1, \dots, J$ et $k = 1, \dots, K_j$) à partir des observations $y_{j,i}$.

2 Modèle hiérarchique bayésien

Les paramètres inconnus pour le problème de segmentation introduit précédemment sont K_j , $l_{j,k}$ et les paramètres de Poisson $\lambda_{j,k}$. Un reparamétrage classique [3, 5] consiste à introduire des indicateurs $r_{j,i}$, $j \in \{1, \dots, J\}$, $i \in \{1, \dots, n\}$ tels que :

$$\begin{cases} r_{j,i} = 1 \text{ s'il y a une rupture à l'instant } i \text{ du } j^{\text{ième}} \text{ signal,} \\ r_{j,i} = 0 \text{ sinon,} \end{cases}$$

avec $r_{j,n} = 1$ (cette condition assure que le nombre de paliers et le nombre de ruptures du $j^{\text{ième}}$ signal sont égaux à $K_j = \sum_{i=1}^n r_{j,i}$). Le vecteur inconnu des paramètres devient alors $\theta = (r, \lambda)$ avec $r = (r_1, \dots, r_n)$, $r_i = (r_{1,i}, \dots, r_{J,i})^T$ et $\lambda = (\lambda_1, \dots, \lambda_J)$, $\lambda_j = (\lambda_{j,1}, \dots, \lambda_{j,K_j})^T$. Il est important de noter que le vecteur de paramètres θ appartient à un espace $\Theta = \{0, 1\}^{nJ} \times \prod_{j=1}^J \mathbb{R}_+^{K_j}$ dont la dimension dépend de la valeur des paramètres K_j . Ce travail se propose d’estimer le vecteur de paramètres θ en utilisant une méthode d’inférence Bayésienne hiérarchique. Pour cela, il est nécessaire de définir la vraisemblance des observations et les lois *a priori* des paramètres inconnus.

2.1 Vraisemblance

La vraisemblance du vecteur des observations y s’écrit :

$$\begin{aligned} f(y|\theta) &= \prod_{j=1}^J \prod_{k=1}^{K_j} \prod_{i \in I_{j,k}} \frac{\lambda_{j,k}^{y_{j,i}} \exp(-\lambda_{j,k})}{y_{j,i}!}, \\ &\propto \prod_{j=1}^J \prod_{k=1}^{K_j} \lambda_{j,k}^{s_{j,k}(r)} \exp(-\lambda_{j,k} n_{j,k}(r)), \end{aligned} \quad (1)$$

où \propto signifie "proportionnel à", $s_{j,k}(r) = \sum_{i \in I_{j,k}} y_{j,i}$ (somme des points du $k^{\text{ième}}$ intervalle $I_{j,k}$ du $j^{\text{ième}}$ signal) et $n_{j,k}(r) = l_{j,k} - l_{j,k-1}$ (nombre de points du segment $I_{j,k}$).

2.2 Lois *a priori* des paramètres

2.2.1 Vecteur des indicateurs

Nous supposons que la probabilité d'avoir $r_i = \epsilon$, ne dépend pas de i , avec $r_i = (r_{1,i}, \dots, r_{J,i})^T$ et $\epsilon = (\epsilon_1, \dots, \epsilon_J)^T \in \{0, 1\}^J$. La loi *a priori* des indicatrices s'écrit alors :

$$f(r|P) = \prod_{\epsilon \in \{0,1\}^J} P_\epsilon^{S_\epsilon(r)},$$

où $P_\epsilon \in \{P_{0\dots 0}, \dots, P_{1\dots 1}\}$ et $S_\epsilon(r)$ est le nombre d'instantanés tels que $r_i = \epsilon$. Par exemple, dans le cas de deux signaux y_1 et y_2 (i.e. $J = 2$), la distribution *a priori* peut s'écrire :

$$f(r|P) = P_{00}^{S_{00}(r)} P_{10}^{S_{10}(r)} P_{01}^{S_{01}(r)} P_{11}^{S_{11}(r)},$$

où $S_{00}(r) = \sum_{i=1}^{n-1} (1 - r_{1i})(1 - r_{2i})$, $S_{11}(r) = \sum_{i=1}^{n-1} r_{1i}r_{2i}$, $S_{10}(r) = \sum_{i=1}^{n-1} r_{1i}(1 - r_{2i})$ et $S_{01}(r) = \sum_{i=1}^{n-1} (1 - r_{1i})r_{2i}$. A une grande valeur de P_ϵ correspondra une forte probabilité d'avoir la configuration $r_i = \epsilon$. Par exemple, en choisissant une grande valeur pour $P_{0\dots 0}$ (resp. pour $P_{1\dots 1}$), nous favoriserons l'absence (resp. la présence) de ruptures simultanées dans toutes les séquences. Ce choix permet d'introduire une corrélation entre les points de ruptures dans les différents signaux temporels.

2.2.2 Paramètres de Poisson

Nous supposons que les paramètres des lois de Poisson suivent une loi Gamma :

$$f(\lambda_{j,k}|\nu, \gamma) \sim \mathcal{G}(\nu, \gamma),$$

où $\nu = 1$ et γ est un hyperparamètre qui peut être ajusté. En supposant l'indépendance des paramètres $\lambda_{j,k}$, la loi *a priori* de λ peut s'écrire :

$$\begin{aligned} f(\lambda|\gamma) &= \prod_{j=1}^J \prod_{k=1}^{K_j} f(\lambda_{j,k}|\nu, \gamma), \\ &= \prod_{j=1}^J \left[\frac{\gamma^{\nu K_j} e^{-\gamma \sum_{k=1}^{K_j} \lambda_{j,k}}}{\Gamma(\nu)^{K_j}} \prod_{k=1}^{K_j} \left(\lambda_{j,k}^{\nu-1} I_{\mathbb{R}^+}(\lambda_{j,k}) \right) \right], \end{aligned}$$

où $I_{\mathbb{R}^+}$ est la fonction indicatrice sur \mathbb{R}^+ ($I_{\mathbb{R}^+}(x) = 1$ si $x \geq 0$ et $I_{\mathbb{R}^+}(x) = 0$ sinon).

Le vecteur des hyperparamètres est noté $\Phi = (P, \gamma)$. Bien sûr, la qualité de la segmentation dépend fortement de la valeur de Φ . Dans le cas de certaines applications, les hyperparamètres peuvent être fixés à partir d'informations disponibles concernant les signaux observés. Cependant, afin d'augmenter la robustesse de l'algorithme, les hyperparamètres peuvent être considérés comme des variables aléatoires dont les lois *a priori* restent vagues, i.e. n'apportent pas d'information, comme c'est le cas dans [4]. Cette stratégie qui implique différents niveaux dans le paradigme bayésien est connue sous le nom d'analyse Bayésienne hiérarchique. Cette analyse impose de définir des lois *a priori* sur les hyperparamètres qui sont détaillées dans la partie suivante.

2.3 Lois *a priori* des hyperparamètres

• **Hyperparamètre γ** : la loi *a priori* pour γ est une loi non-informative de Jeffrey qui traduit l'absence d'information sur cet hyperparamètre :

$$f(\gamma) = \frac{1}{\gamma} I_{\mathbb{R}^+}(\gamma).$$

• **Hyperparamètre P** : la loi *a priori* de P est une loi de Dirichlet de vecteur paramètre $(\alpha_{0\dots 0}, \dots, \alpha_{1\dots 1})^T$. Cette loi est définie sur l'ensemble $S = \left\{ P; \sum_{\epsilon \in \{0,1\}^J} P_\epsilon = 1, P_\epsilon > 0 \right\}$ et sera notée $f(P|\alpha) \sim \mathcal{D}_{2^J}(\alpha)$. Les paramètres γ et P sont supposés indépendants. On peut alors écrire :

$$f(\Phi|\alpha) = \left(\prod_{\epsilon \in \{0,1\}^J} P_\epsilon^{\alpha_\epsilon - 1} \right) \frac{1}{\gamma} I_{\mathbb{R}^+}(\gamma). \quad (2)$$

2.4 Loi *a posteriori* de θ

La loi *a posteriori* du vecteur des paramètres inconnus $\theta = (\lambda, r)$ peut être calculée à partir de la structure hiérarchique suivante :

$$f(\theta|y) = \int f(\theta, \Phi|y) d\Phi \propto \int f(y|\theta) f(\theta|\Phi) f(\Phi) d\Phi,$$

où $f(y|\theta)$ et $f(\Phi)$ ont été définies dans (1) et (2). Des calculs élémentaires permettent d'obtenir le résultat suivant :

$$\frac{f(r, \gamma|y)}{C(r|y)} \propto \frac{1}{\gamma} \prod_{j=1}^J \left[\left(\frac{\gamma^\nu}{\Gamma(\nu)} \right)^{K_j} \prod_{k=1}^{K_j} \frac{\Gamma(n_{j,k}(r) + \nu)}{(s_{j,k}(r) + \gamma)^{n_{j,k}(r) + \nu}} \right], \quad (3)$$

avec

$$C(r|y) = \frac{\prod_{\epsilon \in \{0,1\}^J} \Gamma(S_\epsilon(r) + \alpha_\epsilon)}{\Gamma\left(\sum_{\epsilon \in \{0,1\}^J} (S_\epsilon(r) + \alpha_\epsilon)\right)}$$

où $\Gamma(t)$ est la fonction Gamma. La loi *a posteriori* (3) a une forme trop complexe pour obtenir une expression simple des estimateurs bayésiens (comme l'estimateur minimisant l'erreur quadratique moyenne ou l'estimateur du maximum *a posteriori*). Nous nous proposons de générer des échantillons qui sont asymptotiquement distribués suivant la loi *a posteriori* (3) qui nous intéresse à l'aide d'une méthode MCMC. Ces échantillons seront ensuite utilisés pour estimer les paramètres inconnus du modèle de segmentation conjointe. Nous étudions dans cet article une stratégie faisant appel à un échantillonneur de Gibbs, stratégie qui a également été adoptée dans [4]. Notons cependant les différences suivantes entre les deux approches. La méthode proposée dans cet article ne nécessite pas d'utiliser des sauts réversibles puisque les paramètres inconnus apparaissant dans (3) (obtenus après reparamétrisation) appartiennent à un ensemble de dimension fixe. Ceci simplifie considérablement la complexité calculatoire de l'algorithme. La méthode proposée dans cet article permet également de segmenter conjointement plusieurs signaux, contrairement à [4].

3 Échantillonneur de Gibbs

L'échantillonneur de Gibbs est une méthode de simulation qui consiste à générer des données distribuées suivant une loi d'expression analytique connue à une constante multiplicative près.

Dans cet article, nous utilisons cette méthode pour générer des données distribuées suivant la loi $f(r, \gamma|y)$. La génération se fait en deux étapes : 1) génération d'indicatrices suivant $f(r|\gamma, y)$ et 2) génération d'hyperparamètres γ suivant $f(\gamma|r, y)$.

3.1 Génération suivant $f(r|\gamma, y)$

Cette génération s'effectue à partir de $(n-1)$ tirages suivant la loi des variables $r_i|\gamma, y$ où $r_i = (r_{1,i}, \dots, r_{J,i})$. Ce vecteur aléatoire est discret et prend ses valeurs dans $\mathcal{F} = \{0, 1\}^J$. Ainsi cette loi est totalement caractérisée par les probabilités $P(r_i = \epsilon|\gamma, y)$, $\epsilon \in \mathcal{F}$. En utilisant les notations $r_{p:q} = (r_p, \dots, r_q)$ et $r_{-i} = (r_{1:i-1}, r_{i+1:n})$, on a le résultat suivant :

$$P(r_i = \epsilon|r_{-i}, \gamma, y) \propto f(r_i(\epsilon), \gamma|y),$$

où $r_i(\epsilon) = (r_{1:i-1}, \epsilon, r_{i+1:n})$ et $\epsilon \in \{0, 1\}^J$. On obtient alors une expression simple de $P(r_i = \epsilon|r, y)$ après normalisation.

3.2 Génération suivant $f(\gamma|r, y)$

Pour générer des hyperparamètres de loi $f(\gamma|r, y)$, il est très commode de simuler des vecteurs distribués suivant la loi conjointe $f(\gamma, \lambda|r, y)$ en utilisant des mouvements de Gibbs. Cette étape peut alors se décomposer comme suit :

- **Tirage d'échantillons selon $f(\lambda|\gamma, r, y)$** : en se penchant de près sur la distribution conjointe $f(\theta, \Phi|y)$, on peut obtenir le résultat suivant :

$$\lambda_{j,k}|\gamma, r, y \sim \mathcal{G}(s_{j,k}(r) + \nu, n_{j,k}(r) + \gamma),$$

où $\mathcal{G}(a, b)$ est la loi Gamma de paramètre a et b .

- **Génération suivant $f(\gamma|\lambda, r, y)$** : cette simulation est effectuée comme suit :

$$\gamma|\lambda, r, y \sim \mathcal{G}\left(\nu \sum_{j=1}^J K_j, \sum_{j=1}^J \sum_{k=1}^{K_j} \lambda_{j,k}\right).$$

3.3 Loi *a posteriori* de l'hyperparamètre P

La loi *a posteriori* de P est intéressante puisqu'elle fournit les probabilités *a posteriori* d'avoir des ruptures simultanées ou non dans les différents signaux observés. Quelques calculs supplémentaires permettent de déterminer la loi *a posteriori* de P conditionnellement au vecteur des indicateurs r , aux observations y et aux paramètres $(\alpha_\epsilon)_{\epsilon \in \mathcal{F}}$. Il s'agit d'une loi de Dirichlet d'ordre 2^J et de paramètres $(\alpha_\epsilon + S_\epsilon(r))_{\epsilon \in \mathcal{F}}$:

$$f(P|r, y, \alpha) \sim \mathcal{D}_{2^J}(\alpha_\epsilon + S_\epsilon(r)). \quad (4)$$

4 Simulations

Les simulations présentées ci-dessous ont été obtenues pour $J = 2$ séquences de $n = 120$ échantillons. Les instants de ruptures pour ces deux séquences sont $l_1 = (20, 50, 100, 120)$, $l_2 = (50, 120)$ tandis que les paramètres des lois de Poisson sont $\lambda_1 = (19, 9, 16, 6)$ et $\lambda_2 = (8, 11)$. Les résultats ont été obtenus après avoir moyenné les échantillons fournis par 64 chaînes de Markov. Le nombre total d'itérations pour chaque chaîne est $N_{MC} = 1000$, qui comprend $N_{bi} = 200$ itérations de chauffage. Nous comparons alors les résultats obtenus pour une segmentation non-conjointe et une segmentation conjointe.

Nous nous intéressons tout d'abord aux nombres de ruptures estimés dans chaque séquence. Les algorithmes de segmentation proposés génèrent des échantillons (r^n, γ^n) distribués suivant la loi conjointe *a posteriori* $f(r, \gamma|y)$ qui permettent de choisir le modèle correct. En effet, le nombre de rupture dans chaque séquence peut être estimé par $\hat{K}_j = \sum_{k=1}^n \hat{r}_{j,k}$. Les figures 1 et 2 montrent les histogrammes des nombres de ruptures estimées dans chaque séquence à partir des derniers échantillons des chaînes de Markov, respectivement dans le cas d'une segmentation non conjointe et d'une segmentation conjointe. Ces histogrammes ont des valeurs maximales pour $K_1 = 4$ et $K_2 = 2$ qui correspondent aux nombres réels de changements (on rappelle que par convention il y a une rupture en $n = 120$ pour les deux signaux).

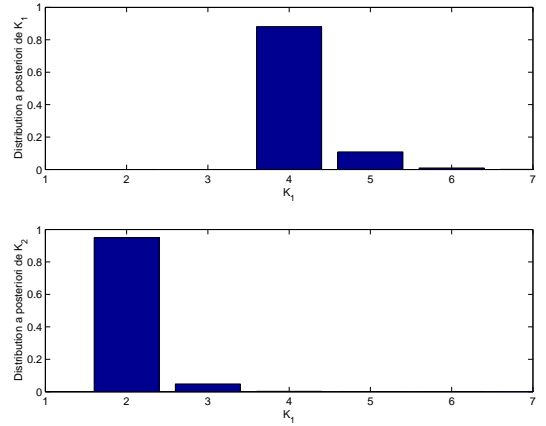


FIG. 1: Lois *a posteriori* des nombres de ruptures (segmentation non-conjointe).

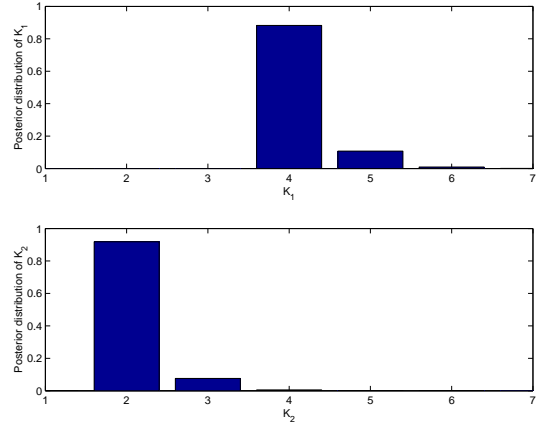


FIG. 2: Lois *a posteriori* des nombres de ruptures (segmentation conjointe).

Les figures 3 et 4 représentent les moyennes des 800 derniers vecteurs d'indicatrices $(r_{1,i}$ et $r_{2,i}$ pour $i = 1, \dots, n$) obtenus dans le cas d'une segmentation conjointe et non conjointe. Notons que puisque $r_{j,i}$ est une variable aléatoire binaire, la moyenne d'un vecteur r distribué suivant la loi $f(r|y)$ fournit les probabilités *a posteriori* d'avoir des ruptures aux instants considérés. Par exemple, il y a une probabilité très importante d'avoir une rupture à la position $l = 100$ pour les deux signaux. Les deux méthodes semblent hésiter quant à la localisation de la première rupture dans la première séquence

($l_{11} = 20$). Enfin, il apparaît clairement que la rupture à la position $l_{12} = l_{21} = 50$ dans la seconde séquence est beaucoup mieux détectée par l'approche conjointe (Fig. 4) que par l'approche non conjointe (Fig. 3).

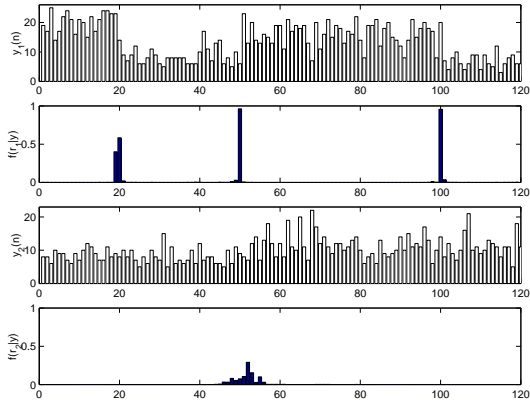


FIG. 3: Lois a posteriori des instants de ruptures (segmentation non-conjointe).

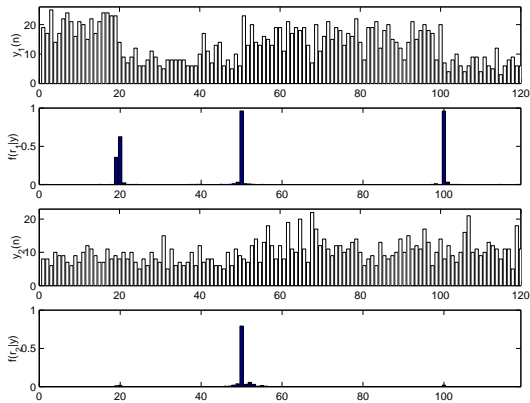


FIG. 4: Distribution a posteriori des instants de ruptures (segmentation conjointe).

L'estimation des paramètres des lois de Poisson est également intéressante puisqu'elle permet la reconstruction des signaux observés. Les lois a posteriori des paramètres $\hat{\lambda}_{1i}$ et $\hat{\lambda}_{2i}$ conditionnellement à $K_1 = 4$ et $K_2 = 2$ sont représentés sur les figures 5 et 6. Ces histogrammes sont en accord avec les valeurs réelles de ces paramètres de Poisson $\lambda_1 = (19, 9, 16, 6)$ et $\lambda_2 = (8, 11)$.

Les derniers résultats illustrent la performance de la méthode pour l'estimation du vecteur d'hyperparamètres P . Les histogrammes des estimateurs \hat{P}_ϵ sont représentés Fig.7. Ces histogrammes coïncident avec les lois a posteriori réelles données par les lois de Dirichlet (4).

5 Conclusion

Cet article présente un algorithme Bayésien hiérarchique permettant de segmenter conjointement plusieurs signaux issus d'une application astrophysique. La méthode proposée est adaptée à des observations Poissonniennes. Les travaux en cours consistent à étendre cette approche à d'autres catégories de données décrites dans [6] et d'effectuer une segmentation conjointe à partir de données issues de capteurs différents (par exemple segmentation de signaux de tension et d'intensité).

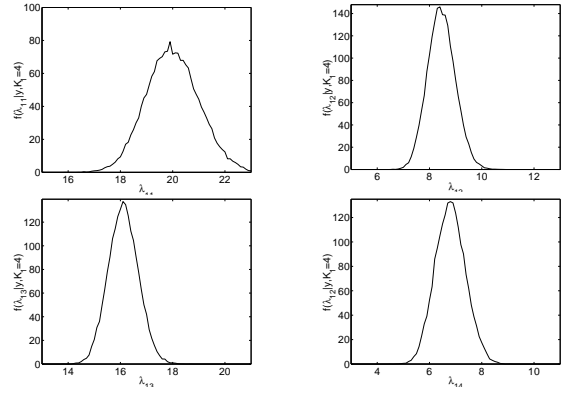


FIG. 5: Lois a posteriori des paramètres λ_{1i} (pour $i = 1, \dots, 4$) conditionnellement à $K_1 = 4$.

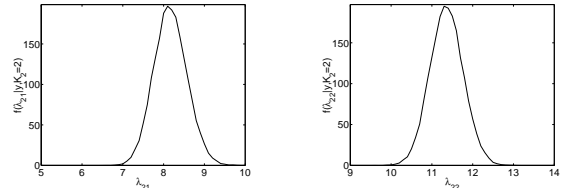


FIG. 6: Lois a posteriori des paramètres λ_{2i} (pour $i = 1, 2$) conditionnellement à $K_2 = 2$.

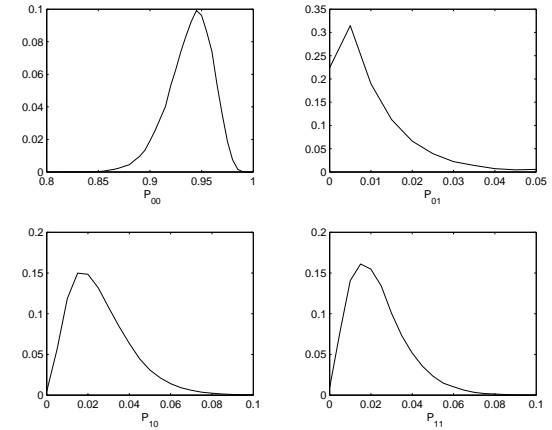


FIG. 7: Lois a posteriori des hyperparamètres P_{00}, P_{01}, P_{10} et P_{11} .

Références

- [1] M. Basseville et I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Englewood Cliffs, Prentice Hall, 1993.
- [2] M. Lavielle, "Optimal segmentation of random processes," *IEEE Trans. Sig. Proc.*, vol. 46, no. 5, pp. 1365-1373, March 1998.
- [3] J-P. Cocquerez et S. Philipp, *Analyse d'images : filtrage et segmentation*. Masson, Sept. 1995.
- [4] E. Punskeya, "Bayesian curve fitting using MCMC with applications to signal segmentation," *IEEE Trans. Sig. Proc.*, vol. 50, no. 3, pp. 747-758, March 2002.
- [5] J.-Y. Tourneret, M. Doisy and M. Lavielle, "Bayesian retrospective detection of multiple changepoints corrupted by multiplicative noise. Application to SAR image edge detection," *Signal Processing*, vol. 83, no. 9, pp. 1871-1887, Sept. 2003.
- [6] J. D. Scargle, "Studies in astronomical time series: v. bayesian blocks, a new method to analyze structure in photon counting data," *The Astrophysical Journal*, pp. 405-418, Sept. 1998.