Classification bayésienne supervisée par processus de Dirichlet

Manuel DAVY[1], Jean-Yves TOURNERET[2]

[1]CNRS/LAGIS et INRIA Futur équipe SequeL,
40 Avenue Halley, 59650 Villeneuve d'Ascq, France

[2]ENSEEIHT/Institut de Recherche en Informatique de Toulouse
2, rue Charles Camichel, BP 7122 - F 31071 Toulouse Cedex 7, France
Manuel.Davy@inria.fr, jean-yves.tourneret@enseeiht.fr

Résumé – La classification supervisée bayésienne de signaux, avec modèle génératif, comporte typiquement deux difficultés. La première est liée à la modélisation statistique des classes (c'est-à-dire à la définition de distributions *a priori* pour les paramètres). La seconde provient des difficultés liées au calcul bayésien. Dans cet article, nous proposons un *a priori* non paramétrique de type «mélange infini de distributions » connu sous le nom de *mélange de processus de Dirichlet*, ainsi qu'un algorithme de calcul bayésien par chaînes de Markov adapté. Ce couple modèle/algorithme est appliqué à la classification de données altimétriques radar.

Abstract – Generative bayesian supervised classification of signals is made difficult by two main issues. The first is related to the definition of a convenient prior distribution for the signal parameter. The second is in computing the quantities needed to perform the classification. In this paper, we propose a nonparametric prior known as the Dirichlet Process Mixture, as well as a suited Monte Carlo Markov Chain algorithm. These model and algorithm are applied to the classification of radar altimetry data.

1 Introduction

Supposons que l'on souhaite classer des signaux (données vectorielles) en K classes. Dans le contexte supervisé, on suppose en outre qu'un ensemble d'apprentissage est disponible, noté $\mathbf{X}_j = \{\mathbf{x}_{1,j}, \ldots, \mathbf{x}_{N_j,j}\}$ pour chaque classe $j = 1, \ldots, K$. Pour chacune des classes, on suppose le modèle génératif des données suivant

$$\mathbf{x}_{i,j} \;=\; F_j(\theta_{i,j}) + \boldsymbol{\epsilon}_{i,j} \tag{1}$$

où le paramètre vectoriels $\theta_{i,j}$ appartient à l'espace Θ_j, $i = 1, \ldots, N_j$ et $\boldsymbol{\epsilon}_{i,j}$ est un bruit aléatoire, supposé additif ici à titre d'exemple. En termes statistiques,

$$\mathbf{x}_{i,j} \;\sim\; \mathsf{L}_j(\mathbf{x}_{i,j}|\theta_{i,j}) \tag{2}$$

où L_j est la vraisembance spécifique à la classe numéro j, pour le paramètre $\theta_{i,j}$. Le but de l'apprentissage bayésien est d'apprendre la distribution de $\theta_{i,j}$ à partir de l'ensemble d'apprentissage $\mathbf{X}_j$, pour chaque classe $j = 1, \ldots, K$.

Approche paramétrique hiérarchique. L'approche paramétrique hiérarchique consiste à définir une distribution *a priori* sur l'espace du paramètre Θ_j, notée $g_j(\theta_j|\boldsymbol{\phi}_j)$, d'hyperparamètre $\boldsymbol{\phi}_j$. La distribution *a posteriori* du paramètre est alors

$$p(\theta_j|\mathbf{X}_j) \;=\; \int g_j(\theta_j|\boldsymbol{\phi}_j) p(\boldsymbol{\phi}_j|\mathbf{X}_j) d\boldsymbol{\phi}_j \tag{3}$$

où, par la règle de Bayes, $p(\boldsymbol{\phi}_j|\mathbf{X}_j) =$

$$\int \ldots \int p(\theta_{1,j}, \ldots, \theta_{N_j,j}, \boldsymbol{\phi}_j|\mathbf{X}_j) d\theta_{1,j} \ldots d\theta_{N_j,j} \tag{4}$$

où $p(\theta_{1,j}, \ldots, \theta_{N_j,j}, \boldsymbol{\phi}_j|\mathbf{X}_j) \;=$

$$h_j(\boldsymbol{\phi}_j) \prod_{i=1}^{N_j} \left[\mathsf{L}_j(\mathbf{x}_{i,j}|\theta_{i,j}) g_j(\theta_{i,j}|\boldsymbol{\phi}_j)\right] \tag{5}$$

On voit que l'apprentissage de la distribution *a posteriori* du paramètre de chaque classe nécessite la définition d'une distribution *a priori* sur l'hyperparamètre, notée $h_j(\boldsymbol{\phi}_j)$. Quand $p(\theta_j|\mathbf{X}_j)$ est apprise, une nouvelle donnée $\mathbf{x}$ est affectée à la classe de plus forte vraisemblance[1] $p(\mathbf{x}|\mathbf{X}_j)$, $j = 1, \ldots, K$, où

$$p(\mathbf{x}|\mathbf{X}_j) \;=\; \int \mathsf{L}_j(\mathbf{x}|\theta_j) p(\theta_j|\mathbf{X}_j) d\theta_j \tag{6}$$

Cette approche peut être mise en œuvre à l'aide de l'algorithme MCMC présenté dans [4]. Un défaut de cette approche est de vouloir modéliser la distribution du paramètre par une distribution *a priori* paramétrique $g_j(\theta_j|\boldsymbol{\phi}_j)$, dont le choix peut être extrêmement arbitraire dans les applications.

Approche proposée. Nous proposons une modélisation plus flexible, plus robuste et numériquement accessible, de la distribution du paramètre. Elle repose sur un modèle de mélange infini, dont la distribution de mélange est un processus de Dirichlet (DP).

2 Processus de Dirichlet et mélanges

Les DP [6, 1] peuvent être vus comme des distributions sur l'espace des distributions. Plus précisément, considérons une distribution de probabilité F_0 définie sur un espace mesurable $(\Theta, \boldsymbol{\Theta})$. Une distribution F sur $(\Theta, \boldsymbol{\Theta})$ est dite distribuée selon un DP (noté $\mathsf{F} \sim \mathcal{DP}(\mathsf{F}; \alpha, \mathsf{F}_0)$) si, pour $A_1, \ldots, A_m \in \boldsymbol{\Theta}$ formant une partition de Θ, on a

$$[\mathsf{F}(A_1), \ldots, \mathsf{F}(A_m)] \;\sim\; \mathcal{D}\big(\cdot; \alpha\mathsf{F}_0(A_1), \ldots, \alpha\mathsf{F}_0(A_m)\big) \tag{7}$$

où $\mathcal{D}$ est la distribution de Dirichlet habituelle. Les DP ont deux paramètres : la distribution de base F_0 et le paramètre $\alpha > 0$.

[1]Nous supposons que toutes les classes ont la même probabilité *a priori*.

Une réalisation F d'un DP est presque sûrement discrète [1], et presque sûrement,

$$\mathsf{F}(d\boldsymbol{\psi}) \;=\; \sum_{k=1}^{\infty} \omega_k \delta_{U_k}(d\boldsymbol{\psi}) \tag{8}$$

La représentation *stick breaking* fournit les poids ω_k et les positions U_k, en répétant pour $k = 1, 2, \ldots$, les étapes suivantes : 1) Echantillonner $U_k \;\sim\; \mathsf{F}_0(\cdot)$ et 2) Echantillonner $\beta_k \;\sim\; \mathcal{B}(1, \alpha)$ puis calculer $\omega_k = \beta_k \prod_{k'=1}^{k-1}(1-\beta_{k'})$, où $\mathcal{B}$ est la distribution beta. On voit que F_0 détermine les positions des composantes discrètes de F, tandis que α règle la variance des poids ω_k. Les DP sont particulièrement adaptés à l'échantillonnage conditionnel. En effet, considérons un ensemble de variables $\{\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_N\}$ dans Ψ i.i.d selon $\mathsf{F}(\cdot)$, où $\mathsf{F} \sim \mathcal{DP}(\mathsf{F}; \alpha, \mathsf{F}_0)$. Alors, pour $i = 1, \ldots, N$,

$$\begin{array}{l} \mathsf{P}(d\boldsymbol{\psi}_i | \boldsymbol{\psi}_{-i}, \mathsf{F}_0, \alpha) \;= \\ \frac{\alpha}{\alpha+N-1}\mathsf{F}_0(d\boldsymbol{\psi}_i) \;+\; \frac{1}{\alpha+N-1}\sum_{i'=1}^{T} {}_{i' \neq i}\, \delta_{\boldsymbol{\psi}_{i'}}(d\boldsymbol{\psi}_i) \end{array} \tag{9}$$

où $\boldsymbol{\psi}_{-i} = \{\boldsymbol{\psi}_{i'}\}_{i'=1,\ldots,N\,,\, i' \neq i}$. Cette formule est connue sous le nom *d'urne de Polya* [2].

Mélanges par Processus de Dirichlet. Une variable aléatoire θ est distribuée selon un mélange par DP (DPM) quand elle est générée ainsi :

1. Echantillonner $\mathsf{F} \sim \mathcal{DP}(\mathsf{F}; \alpha, \mathsf{F}_0)$
2. Echantillonner $\boldsymbol{\psi} \sim \mathsf{F}(\cdot)$
3. Echantillonner $\theta \sim f(\theta|\boldsymbol{\psi})$

La répétition des étapes 2) et 3) génère une famille de variables aléatoires distribuées selon le DPM de densités $f(\cdot|\cdot)$. Chaque itération de l'étape 2) génère un paramètre $\boldsymbol{\psi}_i$ qui détermine le *cluster* auquel θ_i est affecté. Considérant l'urne de Polya, nous remarquons que plusieurs $\boldsymbol{\psi}_i$ partagent la même position dans l'espace des paramètres. Dans la représentation *stick-breaking*, nous voyons que les positions sont U_k avec la probabilité ω_k. Comme pour les mélanges finis, il est possible d'introduire des variables latentes z_t pour chaque θ_i ($i = 1, \ldots, N$). Cela consiste à remplacer l'étape 2) ci-dessus par

2.1) Echantillonner $z \;\sim\; \mathsf{P}(z|\mathsf{F})$ où $\mathsf{P}(z = k|\mathsf{F}) = \omega_k$. Le poids ω_k vient de la représentation *stick-breaking* de F.

2.2) Affecter $\boldsymbol{\psi} \leftarrow U_z$, où U_z vient de la représentation *stick-breaking* de F.

Enfin, en notant le DPM par $\mathsf{G}(d\theta)$ on a :

$$\mathsf{G}(d\theta) \;=\; \int_{\Theta} F(d\theta|\boldsymbol{\psi})\, d\mathsf{F}(\boldsymbol{\psi}) \;\text{ with }\; \mathsf{F} \;\sim\; \mathcal{DP}(\mathsf{F}; \mathsf{F}_0, \alpha) \tag{10}$$

où $F(\cdot|\boldsymbol{\psi})$ est la distribution de densité $f(\cdot|\boldsymbol{\psi})$ sous la mesure dominante.

3 Modéliser la ditribution du paramètre par un DPM

Dans cette section, nous remplaçons la distribution *a priori* du paramètre $\int_{\Phi} g(\theta|\boldsymbol{\phi})h(\boldsymbol{\phi})d\boldsymbol{\phi}$, introduite à la section 1 par le DPM $\mathsf{G}(d\theta)$. Par soucis de clarté des notations, l'indice de classe est omis. Des calculs similaires sont effectués indépendamment dans chaque classe. Nous considérons le modèle hiérarchique, pour chaque classe

$$\begin{array}{rcll} \mathsf{F} & \sim & \mathcal{DP}\big(\mathsf{F}; \mathsf{F}_0, \alpha\big) & (11) \\ \boldsymbol{\psi}_i & \sim & \mathsf{F} & (12) \\ \theta_i & \sim & f(\theta_i|\boldsymbol{\psi}_i) & (13) \\ \mathbf{x}_i & \sim & \mathsf{L}(\mathbf{x}_i|\theta_i) & (14) \end{array}$$

Il est possible d'écrire la distribution *a posteriori* du paramètre (en introduisant les variables latentes $\mathbf{z} = [z_1, \ldots, z_n]$ et les positions des *clusters* $\mathbf{U} = \{U_1, U_2, \ldots\}$),

$$\begin{array}{rcl} p(\theta|\mathbf{X}) & = & \displaystyle\int \sum_{\mathbf{z}} p(\theta, \mathbf{z}, \mathbf{U}|\mathbf{X}) d\mathbf{U} \\ & = & \displaystyle\int \sum_{z,\mathbf{z}} f(\theta|U_z)\mathsf{P}(z|\mathbf{z})\mathsf{P}(d\mathbf{U}, \mathbf{z}|\mathbf{X}) \end{array} \tag{15}$$

où $\mathsf{P}(z|\mathbf{z})$ vérifie $\mathsf{P}(z = k|\mathbf{z}) = \frac{1}{N}\sum_{i=1}^{N} \delta_{k,z_i}$. La distribution *a posteriori* des paramètres suit

$$\begin{array}{l} \mathsf{P}(d\mathbf{U}, \mathbf{z}|\mathbf{X}) = \displaystyle\int_{\Theta^N} \mathsf{P}(d\boldsymbol{\theta}, d\mathbf{U}, \mathbf{z}|\mathbf{X}) \\ \propto \displaystyle\int_{\Theta^N} \mathcal{DP}(\mathsf{F}; \mathsf{F}_0, \alpha) \prod_{i=1}^{N} \mathsf{L}(\mathbf{x}_i|\theta_i) f(\theta_i|U_{z_i}) \end{array} \tag{16}$$

L'inférence est effectuée en appliquant l'algorithme ci-dessous.

Algorithme 1: Apprentissage bayésien par MCMC

% Step 0 : Initialisation
- Pour $i = 1, \ldots, N$, échantillonner $\widetilde{\boldsymbol{\psi}}^{(0)} \sim p(\boldsymbol{\psi}|\widetilde{\boldsymbol{\psi}}_1^{(0)}, \ldots, \widetilde{\boldsymbol{\psi}}_{i-1}^{(0)})$ par l'urne de Polya et en déduire $\widetilde{\mathbf{z}}^{(0)}$ et $\widetilde{\mathbf{U}}^{(0)}$
- Pour $i = 1, \ldots, N$, échantillonner $\widetilde{\theta}_i^{(0)} \;\sim\; f(\theta|\widetilde{\boldsymbol{\psi}}_i^{(0)})$

% Step 1 : Itérations Pour $l = 1, \ldots, L$, faire

1.1- Échantillonner $\widetilde{\mathbf{z}}^{(l)} \sim \mathsf{P}(\mathbf{z}|\widetilde{\mathbf{U}}^{(l-1)}, \widetilde{\boldsymbol{\theta}}^{(l-1)})$ directement (dans le cas où F_0 et $f(\cdot|\cdot)$ sont conjuguée) ou par un pas de Métropolis-Hastings (MH)

1.2- Échantillonner $\widetilde{\mathbf{U}}^{(l)} \sim p(\mathbf{U}|\widetilde{\mathbf{z}}^{(l)}, \widetilde{\boldsymbol{\theta}}^{(l-1)})$, directement ou par un pas de MH

1.3- Échantillonner $\widetilde{\boldsymbol{\theta}}^{(l)} \;\sim\; p(\boldsymbol{\theta}|\widetilde{\mathbf{z}}^{(l)}, \widetilde{\mathbf{U}}^{(l)})$: pour $i = 1, \ldots, N$, échantillonner par un pas de MH $\widetilde{\theta}_i^{(l)} \;\sim\; p(\theta_i|\mathbf{x}_i, \widetilde{U}^{(l)}_{\widetilde{z}_i^{(l)}}) \propto \mathsf{L}(\mathbf{x}_i|\theta_i) f(\theta_i|\widetilde{U}^{(l)}_{\widetilde{z}_i^{(l)}})$

1.4- Échantillonner $\widetilde{\xi}^{(l)} \;\sim\; \mathsf{P}(\xi|\widetilde{\mathbf{z}}^{(l)})$ tel que $\mathsf{P}(\xi = k|\widetilde{\mathbf{z}}^{(l)}) = \frac{1}{N}\sum_{i=1}^{N} \delta_{k,\widetilde{z}_i^{(l)}}$

1.5- Échantillonner $\widetilde{\theta}^{(l)} \sim f(\theta|\widetilde{U}^{(l)}_{\widetilde{\xi}^{(l)}})$

1.6- (Optionel) Echantillonner les hyperparamètres de F_0 et α

Les étapes 1.1 et 1.2 de l'algorithme ci-dessus sont effectuées selon l'une des procédures proposées dans [8], et l'hyperparamètre α peut être mis à jour selon la technique de [9]. Une remarque d'une grande utilité pratique est que le processus de Dirichlet est représenté sous la forme de centroïdes de clusters $\mathbf{U}$ et de variables latentes $\mathbf{z}$. Cette paramétrisation fait qu'aucune troncature de la représentation *stick breaking* n'est réalisée explicitement : seuls les *clusters* actifs sont mis à jour dans l'algorithme (le *clusters* j est actif si au moins un paramètre θ_i y est associé par la variable latente $z_i = j$).

4 Application à la classification de données altimétriques radar

Le satellite Poseïdon est équipé d'un radar altimétrique fournissant des données sous forme de séries temporelles de longueur utile $T = 104$ échantillons, dont 10 sont représentées à la figure 1. Elles peuvent être décrites par le modèle de Hayne [3, 7, 5], valable pour les zones d'océans, avec

$$\mathbf{x}_t = F(t,\theta)\boldsymbol{\epsilon}_t, \quad t = 1,\dots,T \tag{17}$$

où le bruit $\boldsymbol{\epsilon}_t$ est blanc de distribution gamma $\mathcal{G}a(100,100)$ et

$$F(t,\theta) = P_n + \frac{a_\zeta \sigma_0}{2}\left[1+\operatorname{erf}\left(\frac{t-\tau-c_\zeta\sigma_c^2}{\sqrt{2}\sigma_c}\right)\right] \exp\left[-c_\zeta\left(t-\tau-\frac{c_\zeta\sigma_c^2}{2}\right)\right] \tag{18}$$

avec

$$a_\zeta = \exp\left(\frac{-4\sin^2\zeta}{\xi_2}\right) \tag{19}$$

$$c_\zeta = \xi_1\left[\cos(2\zeta) - \frac{\sin^2(2\zeta)}{\xi_2}\right] \tag{20}$$

Le vecteur paramètre inconnu est

$$\theta = [\tau/\tau^{\text{ref}}, \sigma_0/\sigma_0^{\text{ref}}, \text{SWH}/\text{SWH}^{\text{ref}}, \zeta/\zeta^{\text{ref}}] \in \Theta \subset \mathbb{R}^4$$

avec $\tau > 0$ le temps de propagation de l'onde radar, $\sigma_0 > 0$ un coefficient de rétrodiffusion spécifique à la zone d'océan observée, $\text{SWH} = 2c_{\text{light}}\sigma_s$ où $\sigma_s^2 = \sigma_c^2 - \sigma_p^2$ (c_{light} est la vitesse de la lumière et σ_p^2 est connu) est la hauteur des vagues et ζ est l'angle de dépointage de l'antenne. Toutes les autres grandeurs sont connues. La vraisemblance est donc

$$\mathsf{L}(\mathbf{x}|\theta) \propto \exp\left(-L\sum_{t=1}^{T}\left[\log F(t,\theta) + \frac{\mathbf{x}_t}{F(t,\theta)}\right]\right) \tag{21}$$

La distribution de mélange est supposée gaussienne $f(\theta|\boldsymbol{\psi}) = \mathcal{N}(\mu,\Sigma)$ où $\boldsymbol{\psi} = \{\mu,\Sigma\}$ contient le vecteur moyenne μ et la matrice de covariance Σ. Par simplicité, nous supposons que F_0 est une normale inverse Wishart

$$\mathsf{F}_0(\boldsymbol{\psi}) = \mathcal{N}(\mu;\mu_0,\Sigma/\kappa_0)\mathcal{IW}(\Sigma;\nu_0,\Lambda_0), \tag{22}$$

où μ_0, κ_0, ν_0, Λ_0 sont des hyperparamètres.

L'algorithme 1 est mis en œuvre ici, dans une version où les hyperparamètres κ_0, ν_0 et Λ_0 sont pourvus de distributions *a priori* et échantillonnés dans l'étape 1.6 de l'algorithme 1.

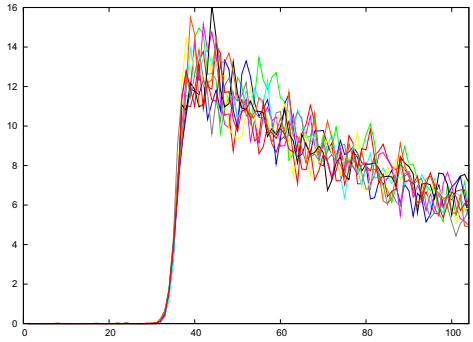

FIG. 1 – Dix signaux radar altimétriques obtenus au dessus d'océans.

Afin de vérifier la capacité de l'algorithme et du modèle à bien apprendre la distribution *a posteriori* du paramètre, nous avons effectué 100 exécutions de l'algorithme 1 pour $N = 100$ signaux d'apprentissage sur 5000 itérations, dont 3000 itérations de chauffage. Pour chaque exécution, les estimées $\widehat{\boldsymbol{\theta}}_i^{\text{MMSE}}$ ont été comparées aux valeurs $\boldsymbol{\theta}_i^{\text{gt}}$ utilisées pour synthétiser les données (pour $i = 1,\dots,100$) via l'indice d'erreur quadratique

$$\text{Err}_{\boldsymbol{\theta}}(N) = \frac{1}{N}\sum_{i=1}^{N}(\widehat{\boldsymbol{\theta}}_i^{\text{MMSE}} - \boldsymbol{\theta}_i^{\text{gt}})^{\mathsf{T}}\mathbf{S}^{-1}(\widehat{\boldsymbol{\theta}}_i^{\text{MMSE}} - \boldsymbol{\theta}_i^{\text{gt}}) \tag{23}$$

où $\mathbf{S}^{-1} = \text{diag}(\tau^{\text{ref}}, \sigma_0^{\text{ref}}, \text{SWH}^{\text{ref}}, \zeta^{\text{ref}})$ et

$$\widehat{\boldsymbol{\theta}}_i^{\text{MMSE}} = (L - L_{\text{burn in}})^{-1}\sum_{l=L_{\text{burn in}}+1}^{L}\widetilde{\boldsymbol{\theta}}_i^{(l)} \tag{24}$$

La moyenne sur 100 exécutions de $\text{Err}_{\boldsymbol{\theta}}(N)$ est de 0.109 avec un écart type de 0.063. La figure 2 montre l'évolution moyenne de l'estimation au cours des itérations.

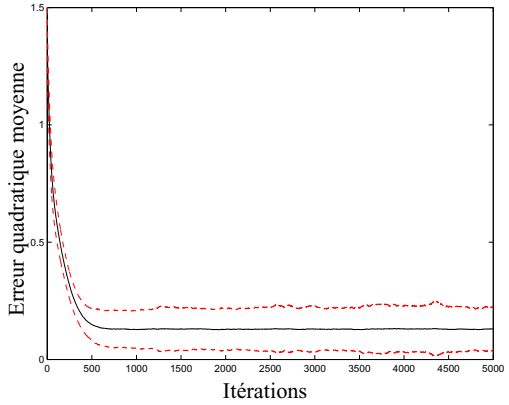

FIG. 2 – Evolution de l'erreur quadratique moyenne du paramètre $\boldsymbol{\theta}_i$, $i = 1,\dots,100$ au cours des 5000 itérations de l'algorithme MCMC, moyenné sur 100 exécutions. Les courbes pointillées représentent l'intervalle de 95% de confiance (± 2 écart-types).

La figure 3 représente les histogrammes des quatre compsantes des échantillons $\widetilde{\theta}^{(l)}$, $l \in \{3001,\dots,5000\}$ pour une exécution typique de l'algorithme 1, ainsi que les densités de probabilité utilisées pour générer les valeurs des paramètres utilisés en simulation. Dans le cas du troisième paramètre, le mélange de processus de Dirichlet a permis de bien apprendre les deux modes de la distribution. La figure 4 représente le même type de résultats, dans le cas où seuls $N = 10$ signaux d'apprentissage sont utilisés. Là encore, on voit que le modèle de DPM apprend bien la distribution du paramètre. L'histogramme de la composante 3 n'est pas bimodal, ce qui montre que le modèle régularise bien lorsque le nombre de données d'apprentissage est faible. Mettre deux modes serait revenu à faire une hypothèse forte compte tenu de la faible information disponible. Enfin, la figure 5 représente ces histogrammes dans le cas où $N = 100$ données d'apprentissage réelles sont utilisées.

5 Discussion, Conclusion et Perspectives

La modélisation proposée permet une grande flexibilité tout en contrôlant la complexité du modèle. Les simulations ont

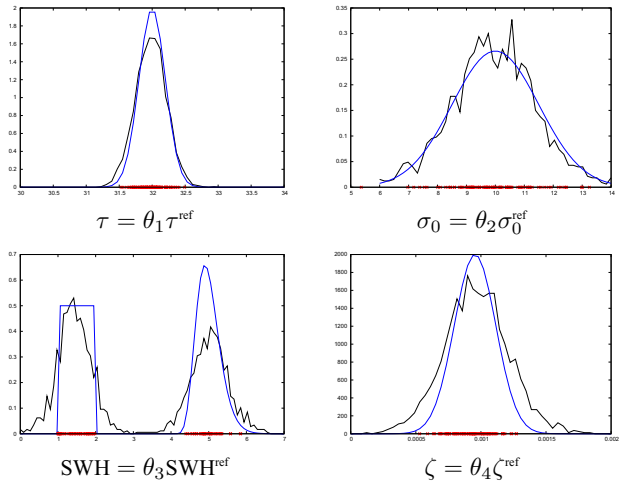

FIG. 3 – En noir, histogrammes des quatre composantes du vecteur paramètre calculés à partir des échantillons $\widetilde{\theta}^{(l)}$ pour $l \in \{3001, \ldots, 5000\}$. En bleu, densités de probabilité utilisées pour tirer les quatre paramètres (représentés par les croix rouges) lors de la synthèse des $N = 100$ signaux.

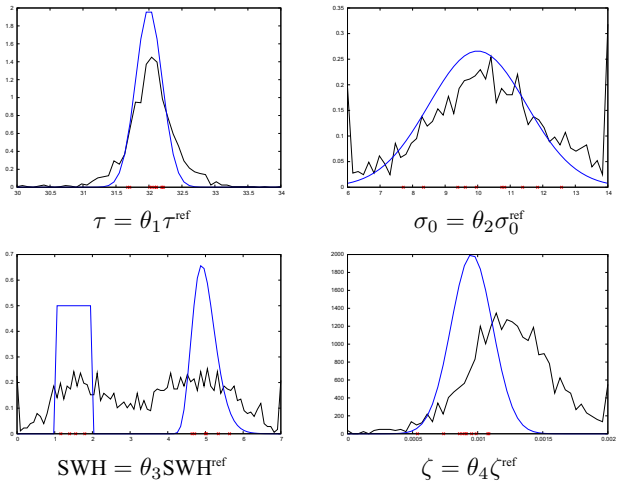

FIG. 4 – Mêmes histogrammes qu'à la figure 3 dans le cas où le nombre de données d'apprentissage est $N = 10$

montré qu'il était possible d'apprendre des distributions multimodales ou plus généralement, non gaussiennes. Le choix de la distribution de base F_0 est peu contraignant en pratique, en particulier du fait que ses paramètres sont estimés aussi. Les résultats de simulation présentés montrent la capacité d'apprentissage du modèle, dont découle directement l'efficacité de classification. L'algorithme MCMC proposé est assez simple à mettre en œuvre, et il peut s'écrire sous forme générique pour la plupart des étapes.

La suite de ce travail consiste à développer un algorithme de type Monte Carlo Séquentiel pour accélérer la convergence de l'algorithme MCMC présenté ici, et de traiter d'autres applications de classification supervisée où un modèle génératif est disponible.

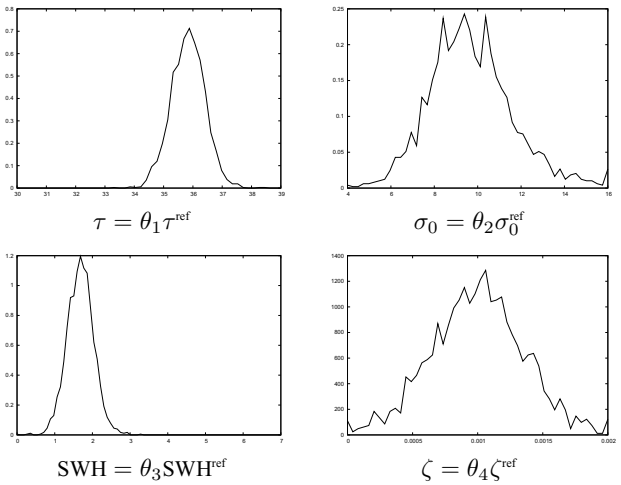

FIG. 5 – Mêmes histogrammes qu'à la figure 3 dans le cas où l'ensemble d'apprentissage est composé de $N = 100$ données réelles.

6 Remerciements

Les auteurs remercient vivement la société CLS et plus particulièrement Laiba Amarouche, Pierre Thibault et Ouan-Zan Zanife pour leurs nombreuses discussions concernant l'altimétrie radar.

Références

[1] C.E. Antoniak. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The annals of statistics*, 2 :1152–1174, 1974.

[2] D. Blackwell and J.B. MacQueen. Ferguson distributions via Polya urn schemes. *The Annals of Statistics*, 1(2) :353–355, 1973.

[3] G. Brown. The average impulse response of a rough surface and its applications. *IEEE Transactions on Antennas and Propagation*, 25(1) :67–74, 1977.

[4] M. Davy, C. Doncarli, and J. Y. Tourneret. Classification of chirp signals using hierarchical Bayesian learning and MCMC methods. *IEEE transactions on Signal Processing*, 50(2) :377–388, February 2002.

[5] J. P. Dumont. ALT-RET-OCE-02 - to perform the ocean-2 retracking - definition, accuracy and specification. Technical report, Collecte Localsation Satellite (CLS), Toulouse, France, October 2001.

[6] T.S. Ferguson. A Bayesian analysis of some nonparametric problems. *The annals of statistics*, 1 :209–230, 1973.

[7] G. Hayne. Radar altimeter mean return waveforms from near-normal-incidence ocean surface scattering. *IEEE Transactions on Antennas and Propagation*, 28(5) :687–692, 1980.

[8] R.M. Neal. Markov chain sampling methods for Dirichlet process mixture models. Technical Report 9815, Dpt of Statistics and department of computer science, University of Toronto, Ontario, Canada, 1998.

[9] M. West. Hyperparameter estimation in Dirichlet process mixture models. Technical report, Institute of Statistics and Decision Sciences, Duke university, 1992.