HAL
open science

Caractérisation de mélange de distributions gaussiennes

Didier Maquin, Gilles Mourot, José Ragot

To cite this version:

Didier Maquin, Gilles Mourot, José Ragot. Caractérisation de mélange de distributions gaussiennes. 6e Congrès International Pluridisciplinaire Qualité et Sûreté de Fonctionnement, Qualita'2005, Mar 2005, Bordeaux, France. pp.CDROM. hal-00965902

HAL Id: hal-00965902

https://hal.science/hal-00965902v1

Submitted on 25 Mar 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CARACTERISATION DE MELANGE DE DISTRIBUTIONS GAUSSIENNES

MAQUIN Didier – MOUROT Gilles – RAGOT José

Institut National Polytechnique de Lorraine
Centre de Recherche en Automatique de Nancy - UMR CNRS 7039
2, avenue de la forêt de Haye
54516 Vandoeuvre cedex – FRANCE
Téléphone : +33 (0)3 83 59 56 83 – Fax : +33 (0)3 83 59 56 44
{Didier.Maquin, Gilles.Mourot, Jose.Ragot}@ensem.inpl-nancy.fr

Résumé : Lorsqu'un processus a un comportement complexe, l'établissement d'un modèle unique pour décrire son évolution globale est difficile. Il peut être alors judicieux de caractériser ce fonctionnement par différents modèles valides dans des zones de fonctionnement limitées. Cette idée, simple, est appliquée ici à la caractérisation de distributions sous forme de somme pondérée de distributions élémentaires, le problème posé étant alors l'identification des caractéristiques propres de chaque distribution et des paramètres de pondération traduisant les poids relatifs des distributions locales.

Abstract: When the behaviour of a process is complex, the obtaining of a unique model describing its global dynamics is difficult. It is therefore judicious to characterize that behaviour using several models that are valid in limited operating regions. This idea, which is very simple, is applied here to the distribution characterization coming from a weighted sum of elementary distributions. The induced problem then concerns the identification of each distribution characteristics and the mixing parameters defining the relative weights of local distributions.

Mots clés : Mélange de distribution, multi-modèle, coefficient de mélange, estimation, distribution gaussienne, modèle local.

Key words: Distribution mixture, multiple model, mixing coefficient, estimation, Gaussian distribution, local model.

1 – Introduction

Il existe de nombreuses situations où la complexité des phénomènes mis en jeu ne peut permettre l'établissement d'un modèle unique traduisant le fonctionnement global d'un système. En revanche, il est souvent plus aisé de caractériser et de décrire ce système dans différentes zones de fonctionnement et de regrouper ensuite ces différentes descriptions à partir de mécanismes de commutation ou de transition. Cette démarche commence à être bien développée et on la rencontre, selon les secteurs scientifiques, sous différents vocables : modèles à commutation [Pearson], mélange de modèles [McLachlan-1], [McLachlan-2], multi-modèles [Gasso], modèles à base de règles, modèles continus par morceaux, modèle à plusieurs régimes, mélange de régression [Wong] ... En statistique également, on est conduit à faire des hypothèses sur la distribution des erreurs de mesure et plus généralement sur la nature aléatoire des variables décrivant un phénomène ; souvent les lois utilisées apparaissent comme des sommes pondérées de lois élémentaires.

Des applications de ce mode de représentation ont été développées dans des domaines variés. Par exemple dans [Santamaria], le mélange de distributions gaussiennes est utilisé pour restaurer un signal sismique inconnu. En traitement d'image [Raja], la reconnaissance de couleurs a été effectuée à partir d'une représentation multi-modale de la distribution des couleurs des pixels. Dans le domaine de l'automatique, la conception de multi-modèles a fait l'objet de nombreux travaux en modélisation [Gasso], commande [Chadli] et diagnostic [Akhenak]. On peut citer également des applications particulières en

analyse de visage [Gross], hydrologie urbaine [Todini], traitement de données spatiales [Govaert] et analyse de texte [Baayen].

Le mélange de modèles résulte donc d'une combinaison de modèles locaux, cette combinaison pouvant être multiplicative ou additive, la seconde solution étant plus répandue pour des raisons de simplicité d'analyse et de synthèse. De plus, la forme additive peut privilégier de manière plus simple la notion de modèle local. En effet, en fonction des poids choisis, la plage de validité d'un modèle local peut être ajustée. Par exemple, un multi-modèle M résultant de l'agrégation de p modèles locaux M_i grâce à des poids μ_i peut décrire de nombreux comportements. Par exemple si $\mu_i = 0, \forall i = 1..p$, $i \neq j$, $\mu_j \neq 0$, alors le modèle global correspond au modèle local M_j. De même, si $\mu_i = 0, \forall i = 1..p$, $i \neq j$, $i \neq k$, $\mu_j \neq 0$, $\mu_k \neq 0$ alors le modèle global traduit un fonctionnement issu d'un compromis ajustable entre les modèles locaux M_j et M_k.

Quelles sont les difficultés inhérentes à l'utilisation des mélanges de modèles ? Outre les aspects propres à tous les problèmes de modélisation liés à l'inconsistance des données qui rendent difficile l'estimation des paramètres des modèles et à l'insuffisance des données qui peuvent rendre non identifiables leurs paramètres, certaines difficultés sont spécifiques aux mélanges de modèles. En premier lieu, citons le compromis à réaliser entre le nombre de modèles locaux et leur complexité. En effet, le but recherché étant la capacité du modèle à traduire correctement le fonctionnement du système à étudier, on peut soit choisir un nombre important de modèles locaux à structure très simple (par exemple linéaire), soit utiliser peu de modèles locaux, chaque modèle local ayant une structure plus complexe. La réponse à ce problème relève de la définition d'objectifs multi-critères spécifiques à chaque problème ; le choix du nombre de modèles locaux reste un problème crucial surtout lorsque les données disponibles sont corrompues par des bruits de mesure [Carreira-Perpinan], [Hawkins]. En second lieu, apparaît la difficulté de la classification des données disponibles pour estimer les paramètres des modèles locaux. En effet, pour estimer les paramètres d'un modèle local, il faut choisir, parmi l'ensemble des données disponibles, celles relatives au fonctionnement dont le modèle local est une image. Finalement, deux problèmes sont à résoudre de façon concomitante : la classification des données, c'est-à-dire leur allocation aux différents modèles locaux, et l'estimation des paramètres des modèles locaux à partir des données ainsi classées.

Dans ce qui suit, ce problème est abordé pour une classe particulière de modèles : les distributions gaussiennes. Dans la section 2, la définition d'un mélange est donnée. La section 3 montre comment estimer les paramètres d'un mélange (cas de deux distributions scalaires et extension à un nombre quelconque de distributions multi-variables) et la procédure est illustrée au cours de la section 4. Quelques perspectives de développement sont ensuite proposées.

2 – Mélange de modèles

Considérons deux variables aléatoires X et Y, à valeurs dans $\mathbf{R}$, de fonctions de répartition respectives F_X et F_Y. Soit maintenant une expérience aléatoire à laquelle on associe la variable aléatoire Z définie de la façon suivante : avec la probabilité p, Z a la loi de X et avec la probabilité $(1 - p)$, Z a la loi de Y. On note $p = \mathsf{P}(Z = X)$ et $1 - p = \mathsf{P}(Z = Y)$. La distribution de Z peut être définie par sa fonction de répartition F_Z. On peut écrire la relation ensembliste : $(Z < t) = ((Z = X) \cup (Z < t)) \cup ((Z = Y) \cup (Z < t))$. Les deux éventualités étant incompatibles, on obtient :

$$F_Z(t) = \mathsf{P}(Z < t) = \mathsf{P}(Z = X \cup Z < t) + \mathsf{P}(Z = Y \cup Z < t) \tag{1a}$$

$$F_Z(t) = \mathsf{P}(Z = X)\mathsf{P}(Z < t \mid Z = X) + \mathsf{P}(Z = Y)\mathsf{P}(Z < t \mid Z = Y) \tag{1b}$$

D'où l'expression de la fonction de répartition :

$$F_Z(t) = pF_X(t) + (1 - p)F_Y(t) \tag{2}$$

On dit que la distribution de Z est un mélange de la distribution de X et de la distribution de Y. On généralise aisément à un mélange de r distributions (r >2). A titre d'illustration, la figure 1 présente le cas d'un mélange de trois distributions normales mono-variables et la figure 2 le cas d'un mélange de deux distributions normales multi-variables.

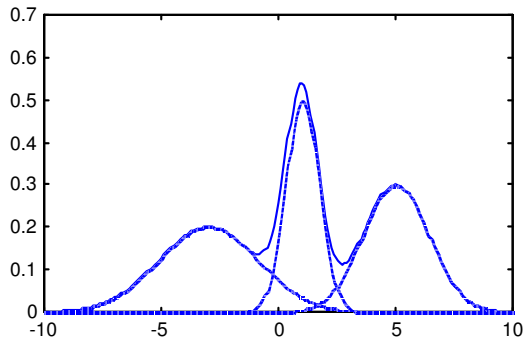

Figure 1 : Mélange de trois gaussiennes mono-variables

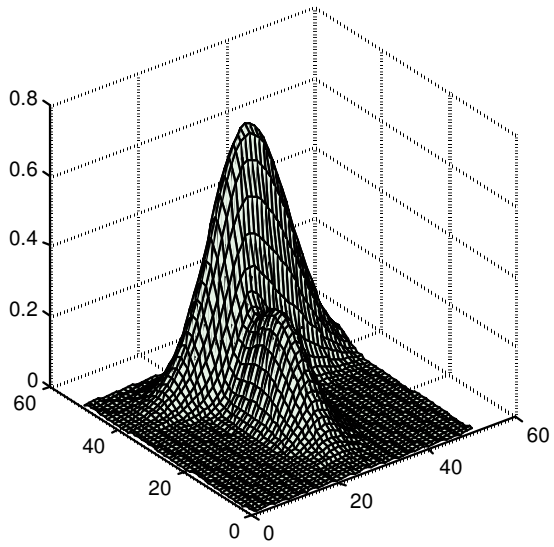

Figure 2 : Mélange de deux gaussiennes muti-variables

3 – Caractérisation du modèle

Dans ce qui suit, on montre comment estimer les paramètres d'un mélange de distributions gaussiennes. Disposant de mesures, l'estimation doit permettre d'identifier les paramètres de chaque distribution locale ainsi que les paramètres de mélange.

3.1 – Modèle à deux distributions mono-variables

On suppose que le système à caractériser comporte deux modes de fonctionnement distincts, chacun étant caractérisé par une distribution gaussienne des mesures. Ces deux modes de fonctionnement sont observés au travers d'un échantillon de N mesures et caractérisés respectivement par des constantes (m_1, σ_1) et (m_2, σ_2). A partir du jeu de mesures disponibles, il convient donc d'estimer ces paramètres, les difficultés étant de savoir à partir de quelles mesures effectuer cette estimation. En effet, à chaque instant, le mode de fonctionnement est inconnu et pour estimer m_1 et σ_1 (respectivement m_2 et σ_2), il faut allouer correctement les mesures à chaque modèle. Le problème est donc double : classer les mesures selon le mode de fonctionnement, estimer les caractéristiques m_1, σ_1, m_2 et σ_2 des modes à partir de cette classification. Bien évidemment les deux tâches sont liées. L'ensemble de la procédure d'estimation est maintenant détaillée.

La loi de distribution de la variable x s'explicite :

$$p(x) = \pi\, p_1(x) + (1-\pi)\, p_2(x) \tag{3}$$

avec les définitions :

$$p_j(x) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{1}{2}\left(\frac{x - m_j}{\sigma_j} \right)^2 \right), \qquad j = 1,2 \tag{4}$$

Pour l'ensemble des N mesures x_i disponibles, avec l'hypothèse d'indépendance des erreurs de mesure, on définit la fonction de log-vraisemblance de l'échantillon :

$$\upsilon = \sum_{i=1}^{N} \log(\pi p_1(x_i) + (1-\pi) p_2(x_i)) \tag{5}$$

où π désigne le paramètre de mélange des deux distributions. Les équations d'optimalité de υ par rapport à m_1, σ_1, m_2 et σ_2 et π s'écrivent :

$$\frac{\partial \upsilon}{\partial m_1} = \sum_{i=1}^{N} \pi \frac{p_{1,i}}{p_i \sigma_1^2}(x_i - m_1) = 0 \tag{6a}$$

$$\frac{\partial \upsilon}{\partial m_2} = \sum_{i=1}^{N} (1-\pi) \frac{p_{2,i}}{p_i \sigma_2^2}(x_i - m_2) = 0 \tag{6b}$$

$$\frac{\partial \upsilon}{\partial \pi} = \sum_{i=1}^{N} \frac{p_{1,i} - p_{2,i}}{p_i} = 0 \tag{6c}$$

$$\frac{\partial \upsilon}{\partial \sigma_1} = \pi \sum_{i=1}^{N} \frac{p_{1,i}}{p_i}\left(\frac{(x_i - m_1)^2}{\sigma_1^3} - \frac{1}{\sigma_1} \right) = 0 \tag{6d}$$

$$\frac{\partial \upsilon}{\partial \sigma_2} = (1-\pi) \sum_{i=1}^{N} \frac{p_{2,i}}{p_i}\left(\frac{(x_i - m_2)^2}{\sigma_2^3} - \frac{1}{\sigma_2} \right) = 0 \tag{6e}$$

avec les notations :

$$p_{1,i} = p_1(x_i), \quad p_{2,i} = p_2(x_i), \quad p_i = \pi p_{1,i} + (1-\pi) p_{2,i} \tag{7}$$

Après simplification d'usage, on déduit de (6)

$$m_1 = \sum_{i=1}^{N} \frac{p_{1,i}}{p_i} x_i \Big/ \sum_{i=1}^{N} \frac{p_{1,i}}{p_i} \tag{8a}$$

$$m_2 = \sum_{i=1}^{N} \frac{p_{2,i}}{p_i} x_i \Big/ \sum_{i=1}^{N} \frac{p_{2,i}}{p_i} \tag{8b}$$

$$\sigma_1^2 = \sum_{i=1}^{N} \frac{p_{1,i}}{p_i}(x_i - m_1)^2 \Big/ \sum_{i=1}^{N} \frac{p_{1,i}}{p_i} \tag{8c}$$

$$\sigma_2^2 = \sum_{i=1}^{N} \frac{p_{2,i}}{p_i}(x_i - m_2)^2 \Big/ \sum_{i=1}^{N} \frac{p_{2,i}}{p_i} \tag{8d}$$

$$\sum_{i=1}^{N} \frac{1}{\pi + \dfrac{p_{2,i}}{p_{1,i} - p_{2,i}}} = 0 \tag{8e}$$

Pour apporter une dernière simplification, on peut noter que :

$$\pi m_1 + (1-\pi) m_2 = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{9a}$$

$$\sum_{i=1}^{N} \frac{p_{1,i}}{p_i} = N \tag{9b}$$

$$\sum_{i=1}^{N} \frac{p_{2,i}}{p_i} = N \tag{9c}$$

Le système (8) est non linéaire en les paramètres m_1, σ_1, m_2, σ_2 et π car $p_{i,1}$ et $p_{i,2}$ dépendent également de ces paramètres. Il n'est pas possible d'exhiber une expression analytique de la solution, mais une procédure itérative de recherche peut être mise en œuvre. D'un point de vue numérique, la résolution peut s'effectuer de façon hiérarchique. En effet, les 4 premières équations du système (8) font directement apparaître une itération sur les paramètres m et σ ; la dernière équation est destinée à fournir le paramètre de mélange et sa résolution est faite selon un algorithme de type gradient. Les estimées à l'itération $(k+1)$ sont fournies sous la forme :

$$m_1^{(k+1)} = \frac{1}{N} \sum_{i=1}^{N} \frac{p_{1,i}^{(k)}}{p_i^{(k)}} x_i \tag{10a}$$

$$m_2^{(k+1)} = \frac{1}{N} \sum_{i=1}^{N} \frac{p_{2,i}^{(k)}}{p_i^{(k)}} x_i \tag{10b}$$

$$\left(\sigma_1^2\right)^{(k+1)} = \frac{1}{N}\sum_{i=1}^{N}\frac{p_{1,i}^{(k)}}{p_i^{(k)}}(x_i - m_1^{(k)})^2 \qquad (10c) \qquad \left(\sigma_2^2\right)^{(k+1)} = \frac{1}{N}\sum_{i=1}^{N}\frac{p_{2,i}^{(k)}}{p_i^{(k)}}(x_i - m_2^{(k)})^2 \qquad (10d)$$

$$\pi^{(k+1)} = \pi^{(k)} + \sum_{i=1}^{N}\frac{1}{\pi^{(k)} + \dfrac{p_{2,i}^{(k)}}{p_{1,i}^{(k)} - p_{2,i}^{(k)}}} \Bigg/ \sum_{i=1}^{N}\frac{1}{\left(\pi^{(k)} + \dfrac{p_{2,i}^{(k)}}{p_{1,i}^{(k)} - p_{2,i}^{(k)}}\right)^2} \qquad (10e)$$

où $p_{1,i}^{(k)}$, $p_{2,i}^{(k)}$ et $p_i^{(k)}$ sont définis en (3) et (4), dans lesquels les paramètres sont remplacés par leurs valeurs respectives obtenues à l'itération (k). Il faut noter, comme dans toute méthode itérative de point fixe, que la convergence de l'algorithme n'est pas garantie de façon globale et dépend en particulier des valeurs initiales choisies [Carreira-Perpinan] ; la possibilité d'existence de plusieurs points fixes est à considérer. Cependant, en dépit de cette difficulté, l'algorithme itératif de type point fixe reste plus simple à mettre en œuvre que celui du gradient, même si le nombre d'itérations pour atteindre l'optimum est en général plus important. Des améliorations peuvent être apportées, notamment en utilisant un filtre discret afin de réduire les fluctuations des amplitudes des estimées entre deux itérations.

L'intérêt de l'approche proposée est de résoudre globalement toutes les équations d'optimalité. Par comparaison avec l'approche EM [Bilmes], [McLachlan], aucun paramètre caché n'a été introduit.

3.1 – Modèle à r distributions multi-variables

Les densités de probabilité s'appliquent maintenant à une variable vectorielle $x \in \mathbf{R}^n$. Chaque distribution élémentaire j est alors caractérisée par une moyenne $m_j \in \mathbf{R}^n$ inconnue et une matrice de variance $V_j \in \mathbf{R}^{n\times n}$ inconnue. La distribution mélange est la somme pondérée des distributions locales :

$$p(x) = \sum_{j=1}^{r}\pi_j p_j(x) \qquad (11)$$

avec :

$$p_j(x) = \pi_j \frac{1}{\sqrt{2\pi}\left|V_j\right|^{n/2}}\exp\left(-\frac{1}{2}(x-m_j)^T V_j^{-1}(x-m_j)\right) \qquad (12)$$

sous la contrainte de normalisation :

$$p(x) = \sum_{j=1}^{r}\pi_j = 1 \qquad (13)$$

A partir d'un échantillon $x_i, i = 1..N$, on souhaite caractériser la distribution mélange, c'est-à-dire estimer les paramètres locaux m_j et V_j et les paramètres de mélange π_j (dont le nombre est fixé a priori). Le principe de l'estimation est le même que précédemment. A partir de l'échantillon de mesures, on construit la fonction de log-vraisemblance qui généralise l'expression (5). Puis on maximise cette fonction de vraisemblance par rapport à m_j, V_j et π_j, $j = 1..r$ Les équations d'optimalité qui s'en déduisent généralisent les expressions (6) obtenues pour un mélange de deux distributions scalaires.

4 – Exemple

4.1 – Mélange de deux distributions mono-variables

Les caractéristiques théoriques des deux distributions sont respectivement $m_1 = 3$, $\sigma_1 = 1$ et $m_1 = 5$, $\sigma_2 = 1$. Des échantillons comportant 1000 et 500 mesures ont été générés à partir de ces deux

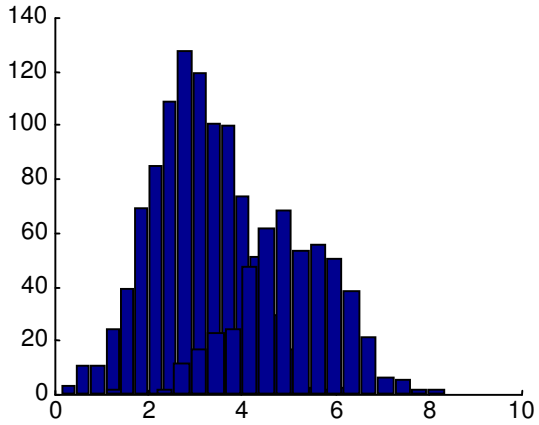

Figure 3 : Histogramme des deux distributions normales mono-variables

distributions ; l'histogramme correspondant est représenté à la figure 3. Les valeurs obtenues, après convergence de l'algorithme développé à la section 3.1, sont respectivement (2.930, 0.956) et (5.073, 1.113). Le coefficient de mélange a été estimé à 0.661 et traduit bien les proportions de taille de la population des deux distributions (ce rapport vaut 1000/1500). La figure 4 montre les distributions théoriques et identifiées.

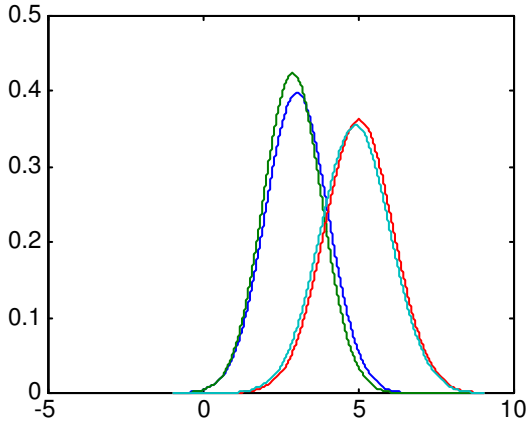

Figure 4 : Distributions théoriques et identifiées

4.2 – Mélange de quatre distributions bivariables

Cet exemple illustre le cas d'un mélange de 4 distributions gaussiennes bivariables ($x \in \mathbf{R}^2$). Les caractéristiques de ces distributions sont résumées dans la table 1. Les données générées peuvent être classées dans un ordre arbitraire, mais, pour des raisons de commodité de représentation, elles sont classées par groupe. La figure 5 indique la répartition des mesures dans le plan de la variable x, les croix indiquant les centres des distributions. On distingue nettement les 4 groupes. La figure 7 donne l'allure des quantités $p_i(x_i), i = 1..4$ qui témoigne de l'appartenance des mesures aux 4 groupes. On retrouve assez nettement la segmentation et les effectifs de classes indiqués dans la table 1. La table 2 fournit les caractéristiques des gaussiennes identifiées, c'est-à-dire les coordonnées des centres, les écarts-types et les effectifs que l'on peut comparer aux valeurs théoriques. La figure 6 montre en superposition les données originales et les valeurs situées à un écart-type du centre des distributions théoriques et identifiées.

Tableau 1 : Caractéristiques vraies des distributions

Distribution	Centre		Ecart-type		Effectif
1	-1.0	2.0	1.50	1.50	200
2	3.0	7.0	1.50	0.50	150
3	9.0	2.0	2.50	2.50	150
4	4.0	-4.0	1.50	4.50	150

Tableau 2 : Caractéristiques estimées des distributions

Distribution	Centre		Ecart-type		Proportion
1	-1.08	2.03	1.53	1.50	29.8
2	3.16	6.89	1.43	0.58	23.6
3	9.19	1.34	2.53	2.37	23.6
4	3.87	-3.99	1.55	4.06	23.0

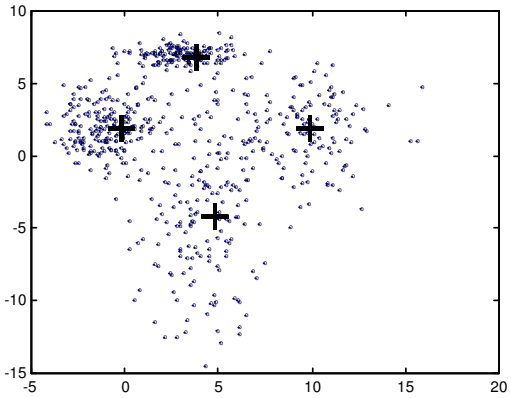

Figure 5 : Fonctions poids et allocation des données aux modèles locaux

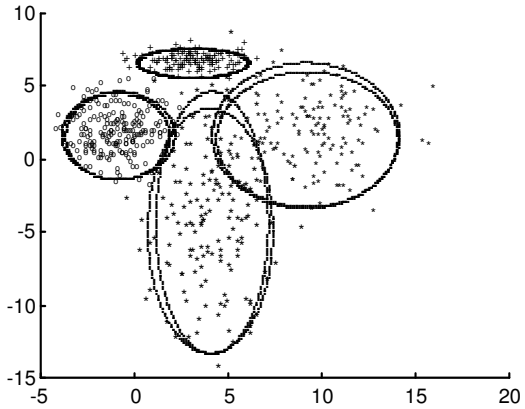

Figure 6 : Données et modèles identifiés

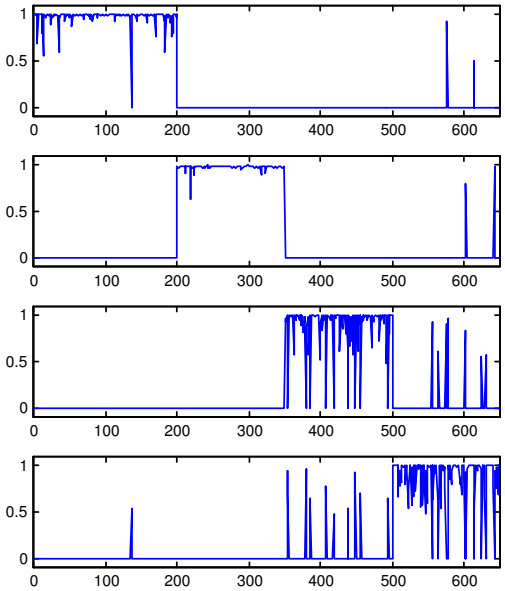

Figure 7 : Fonctions poids et allocation des données aux modèles locaux

5 – Conclusion

En formulant la recherche des paramètres d'un mélange sous forme d'un problème d'optimisation d'une fonction de coût, on a proposé un algorithme résolvant de façon conjointe le problème d'estimation des caractéristiques des modèles locaux et celui d'estimation des paramètres de mélange. Un certain nombre de problèmes devront être examinés par la suite et en particulier celui de la détermination du nombre de distributions locales. Notons de plus que les résultats d'affectation présentés, et notamment dans le deuxième exemple, sont sous forme continue, de telle sorte qu'une mesure peut être affectée à plusieurs distributions locales. Une autre extension, pour laquelle nous venons d'obtenir quelques résultats encourageants, concernent l'affectation binaire des mesures aux distributions.

Références :

[Akhenak] Akhenak A., Chadli M., Ragot J., Maquin D. *Estimation of state and unknown inputs of a non linear system represented by a multiple model*. 11th IFAC Symposium on Automation in Mining, Mineral and Metal processing, MMM 2004, Nancy, France, September 8-10, 2004.

[Baayen] Baayen R.H., Tweedie F.J. *Mixture models and word frequency distributions*. Joint International Conference, Debrecen, Hungary, July 5-10, 1998.

[Bilmes] Bilmes J.A. *A gentle tutorial of the EM algorithm and its application to parameter estimation for gaussian mixture and Hidden Markov models*. Report TR-97-021, International Computer Science Institute, Berkeley, 1998.

[Carreira-Perpinan] Carreira-Perpinan M.A. *Mode-finding for mixtures of Gaussian distributions*. IEEE Trans. on Pattern Analysis and Machine Intelligence 22 (11), pp. 1318-1323, 2000.

[Chadli] Chadli M., Maquin D., Ragot J. *Multiple observers for discrete-time multiple models*. 5th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes, Safeprocess'2003, Washington, D.C., USA, June 9-11, 2003.

[Gasso] Gasso K., Mourot G., Ragot J. *Structure identification of multiple models with output error local models*. 15th IFAC World Congress on Automatic Control, Barcelona, Spain, July 21-26, 2002.

[Govaert] Govaert G. *Classification et modèles de mélange - application aux données spatiales*. Revue Internationale de Géomatique, European Journal of GIS and Spatial Analysis, 9 (4), pp. 457-470, 2000.

[Gross] Gross R., Yang J., Waibel A. *Growing Gaussian mixture models for pose invariant face recognition*. 15th International Conference on Pattern Recognition, Barcelone, Spain, 1, September 3-8, 2000.

[Hawkins] Hawkins D. S., Allen D.M., Stromberg A.J. *Determining the number of components in mixtures of linear models*. Computational Statistics and Data Analysis, 38 (1), pp. 15-48, 2001.

[McLachlan-1] McLachlan G.J., Basford K.E. *Mixture models: inference and applications to clustering*. Marcel Dekker Inc., New York, 1988.

[McLachlan-2] McLachlan G.J., *Finite mixture models*. Wiley Series in Probability and statistics. Wiley Interscience, 2000.

[Pearson] Pearson D.W., Ragot J. *Identifying parameters of local switching models: a geometrical approach*. 11th International Workshop on Systems, Signals and Image Processing, IWSSIP'04, Poznan, Poland, 2004.

[Raja] Raja Y., McKenna S., Gong S. *Segmentation and tracking using colour mixture models*. 3rd Asian Conference on Computer Vision, LACCV'98, Hong Kong, January 8-10, 1998.

[Santamaria] Santamaria I., Pantaleon C.J., Ibanez J., Artés A. *Deconvolution of seismic data using adaptive gaussian mixture*. IEEE Transations on Geoscience and Remote Sensing, 37 (2), pp. 855-859, 1999.

[Todini] Todini E., di Bacco M. *A combined Polya process and mixture distribution approach to rainfall modelling*. Hydrology and Earth System Sciences, 1, pp. 367-378, 1997.

[Wong] Wong C.S., Li W.K. *On a mixture autoregressive model*. Journal of Royal Statistical Society B, 62, pp. 95-115, 2000.