Une version de type recuit simule de l'algorithme EM

Gilles Celeux, Jean Diebolt

▶ **To cite this version:**

Gilles Celeux, Jean Diebolt. Une version de type recuit simule de l'algorithme EM. RR-1123, INRIA. 1989. inria-00075436

HAL Id: inria-00075436
https://inria.hal.science/inria-00075436

Submitted on 24 May 2006

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.

INRIA

UNITE DE RECHERCHE
INRIA-ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
BP 105
78153 Le Chesnay Cedex
France
Tél. (1) 39 63 55 11

Rapports de Recherche

N° 1123

Programme 5
Automatique, Productique,
Traitement du Signal et des Données

UNE VERSION DE TYPE RECUIT SIMULE DE L'ALGORITHME EM

Gilles CELEUX
Jean DIEBOLT

Novembre 1989

RR_1123

1. Introduction

L'algorithme EM (Dempster, Laird et Rubin (1977)) est très répandu pour identifier, par le maximum de vraisemblance, des modèles où certaines données sont manquantes. Néanmoins, il présente des défauts bien connus (lenteur éventuelle, importance souvent cruciale de sa position initiale, convergence vers un col de la vraisemblance). Récemment, nous avons développé un algorithme d'apprentissage probabiliste, l'algorithme SEM (voir par exemple Celeux et Diebolt (1987)), qui répond, en général, aux limitations de l'algorithme EM. Nous présentons un nouvel algorithme, dérivé de l'algorithme SEM, qui peut être considéré comme une version de type recuit simulé de l'algorithme EM.

Cet algorithme, que nous désignerons sous le nom d'algorithme SAEM (Simulated Annealing EM), occupe une place intermédiaire entre les algorithmes EM et SEM. Par rapport à l'algorithme SEM, son intérêt réside dans une plus grande simplicité théorique et une meilleure fiabilité pour de petits échantillons. Par rapport à l'algorithme EM, il vise, comme l'algorithme SEM, à répondre à ses principales limitations.

La section 2 est consacrée à la présentation des algorithmes EM et SEM, et à l'énoncé de leurs principales propriétés. La section 3 est consacrée à la présentation du principe de l'algorithme SAEM dans un cadre très général. Dans la section 4, nous étudions les propriétés de cet algorithme pour l'identification d'un mélange fini de densités de probabilité ; nous donnons, en particulier, un théorème décrivant la convergence presque sûre de l'algorithme SAEM vers un des points fixes de l'algorithme EM, lorsque les composants du mélange ont des densités de probabilité appartenant à la famille exponentielle. La comparaison pratique, sur un mélange gaussien, des trois algorithmes fait l'objet de la section 5.

2. Les algorithmes EM et SEM

2.1. Le contexte statistique

Les deux algorithme s'appliquent à des modèles où l'échantillon observé $\mathbf{x}$, appartenant à un espace $\mathbf{X}$, provient par une application surjective π d'un échantillon inaccessible $\mathbf{y}$, appartenant à un espace $\mathbf{Y}$. On suppose que les espaces $\mathbf{X}$ et $\mathbf{Y}$ sont dotés de mesures σ-finies μ et ν, et que les échantillons $\mathbf{x}$ et $\mathbf{y}$ sont issus de lois admettant, par rapport à ces mesures, des densités, dépendant d'un même paramètre ϕ, notées respectivement $f(\mathbf{x};\phi)$ et $g(\mathbf{y};\phi)$. Du fait de la relation $\mathbf{x} = \pi(\mathbf{y})$, on a informellement

$$f(\mathbf{x};\phi) = \int_{\pi^{-1}(\mathbf{x})} g(\mathbf{y};\phi)\, d\mathbf{y} \qquad (1)$$

où $d\mathbf{y}$ désigne la mesure $\nu(d\mathbf{y})$.

Remarque : nous avons adopté ici la présentation de Dempster et al. (1977). En fait, pour tous les modèles où l'algorithme EM s'applique, on peut écrire $\mathbf{y} = (\mathbf{x},\mathbf{z})$, $\mathbf{z}$ étant l'échantillon des valeurs manquantes $\mathbf{y}$, $\mathbf{z}$ appartenant à un espace $\mathbf{Z}$; on a donc $\mathbf{Y} = \mathbf{X} \times \mathbf{Z}$. L'échantillon est désigné comme l'échantillon complet. L'application π est la projection de $\mathbf{Y}$ sur $\mathbf{X}$ et la formule (1) peut s'écrire

$$f(\mathbf{x},\phi) = \int_{\mathbf{Z}} g(\mathbf{y};\phi)\, d\mathbf{y} \qquad (1\text{bis})$$

Cette décomposition de l'échantillon complet $\mathbf{y}$ en échantillon observé $\mathbf{x}$ et en échantillon manquant $\mathbf{z}$ nous permettra de clarifier l'exposition des algorithmes SEM et SAEM. Le but de l'algorithme EM est de trouver la valeur du paramètre ϕ qui maximise la vraisemblance $f(\mathbf{x};\phi)$

UNE VERSION DE TYPE RECUIT SIMULE DE L'ALGORITHME EM

A SIMULATED ANNEALING TYPE EM ALGORITHM

Gilles CELEUX
INRIA, Rocquencourt

Jean DIEBOLT
CNRS, Paris 6

RESUME

L'algorithme EM est très répandu pour l'estimation par le maximum de vraisemblance de paramètres de modèles où les données sont incomplètes. Nous présentons une version de type recuit simulé de l'algorithme EM. Cet algorithme, désigné algorithme SAEM, est une adaptation de l'algorithme stochastique SEM que nous avons précédemment développé. Comme ce dernier, l'algorithme SAEM répond aux limitations bien connues de l'algorithme EM ; mais de plus il se comporte mieux pour traiter de petits échantillons. Par ailleurs, il est plus simple à appréhender que l'algorithme SEM dans la mesure où il converge presque sûrement tandis que l'algorithme SEM converge en loi. Ici, on limite la présentation détaillée de l'algorithme SAEM au problème des mélanges de lois de probabilité. On établit un théorème qui assure que toute suite d'estimés par SAEM converge p.s. vers un maximum local de la fonction de vraisemblance. On conclut cet article par une étude comparative du comportement pratique des 3 algorithmes EM, SEM et SAEM.

Mots-clés : maximum de vraisemblance, algorithme stochastique, mélange, recuit simulé.

SUMMARY

The EM algorithm is a widely-applicable approach for computing maximum likelihood estimates for incomplete data. We present a simulated annealing type EM algorithm : SAEM. This algorithm is an adaptation of the stochastic EM algorithm (SEM) that we have previously developed. Like SEM, SAEM overcomes most of the well-known limitations of EM. Moreover, SAEM performs better for small samples. Furthermore, SAEM appears to be more tractable than SEM, since it provides almost sure (a.s.) convergence, while SEM provides convergence in distribution. Here, we focus on the mixture problem. We state a theorem which asserts that each SAEM sequence converges a.s. to a local maximizer of the likelihood function. We close this paper with a comparative study, based on numerical simulations, of these three algorithms.

Keywords : maximum likelihood, mixture, simulated annealing, stochastic algorithm.

PAPIER RÉCUPÉRÉ ET RECYCLÉ

en tirant parti de la relation (1) qui lie la vraisemblance de l'échantillon incomplet $\mathbf{x}$ à la vraisemblance $\mathbf{g}(\mathbf{y};\phi)$ de l'échantillon complet, mais non observé, $\mathbf{y}$. En effet, dans la plupart des cas, il est beaucoup plus facile de maximiser $\mathbf{g}(\mathbf{y};\phi)$ que de maximiser $\mathbf{f}(\mathbf{x};\phi)$.

2.2. L'algorithme EM

L'idée de base de cet algorithme consiste à remplacer la vraisemblance des données complètes, incalculable puisque $\mathbf{y}$ n'est pas observé, par son espérance conditionnelle sachant $\mathbf{x}$ et une valeur courante ϕ^n du paramètre.

Soit $\mathbf{k}(\mathbf{y}|\mathbf{x};\phi) = \mathbf{g}(\mathbf{y};\phi)/\mathbf{f}(\mathbf{x};\phi)$ la densité conditionnelle de $\mathbf{y}$ sachant $\mathbf{x}$ et ϕ ; soit $L(\phi) = \ell n\ \mathbf{f}(\mathbf{x};\phi)$. On a

$$L(\phi) = \ell n\ \mathbf{g}(\mathbf{y};\phi) - \ell n\ \mathbf{k}(\mathbf{y}|\mathbf{x};\phi) \qquad (2)$$

On introduit

$$Q(\phi';\phi) = E\left\{\ell n\ \mathbf{g}(\mathbf{y};\phi')|\mathbf{x};\phi\right\} = \int_{\pi^{-1}(\mathbf{x})} \mathbf{k}(\mathbf{y}|\mathbf{x};\phi)\ \ell n\ \mathbf{g}(\mathbf{y};\phi)\ d\mathbf{y}$$

que l'on suppose exister pour tout couple (ϕ',ϕ).

Partant d'une position ϕ^0, une itération de l'algorithme EM, qui à ϕ^n associe $\phi^{n+1} = T(\phi^n)$, est définie ainsi :

étape E : calcul de $Q(\phi;\phi^n)$; cette étape revient à calculer la densité conditionnelle $\mathbf{k}(\mathbf{y}|\mathbf{x};\phi^n)$;
étape M : choix de ϕ^{n+1} qui maximise $Q(\phi;\phi^n)$.

La caractéristique principale de l'algorithme EM est la suivante : $L(T(\phi)) \geq L(\phi)$ pour tout ϕ, l'égalité ayant lieu si et seulement si $Q(T(\phi);\phi) = Q(\phi;\phi)$ et $\mathbf{k}(\mathbf{y}|\mathbf{x};T(\phi)) = \mathbf{k}(\mathbf{y}|\mathbf{x};\phi)$ presque sûrement. Il s'ensuit que toute suite (ϕ^n) engendrée par l'algorithme EM vérifie l'inégalité $L(\phi^{n+1}) \geq L(\phi^n)$.

Une étude détaillée de la convergence de l'algorithme EM a été réalisée par Wu (1983). Sous des conditions assez générales et vérifiées dans beaucoup de situations pratiques, il a montré que la suite $(L(\phi^n))$ convergeait vers une valeur stationnaire de la vraisemblance ; mais, en général, on ne peut pas affirmer que l'algorithme EM converge vers un maximum local de la vraisemblance, à moins de supposer des conditons très contraignantes.

Indiquons, d'autre part, une propriété que nous utiliserons dans la section 4. Notons

$$H(\phi';\phi) = E\left\{\ell n\ \mathbf{k}(\mathbf{y}|\mathbf{x};\phi')|\mathbf{x};\phi\right\}$$

Remarquons que (Dempster et al. (1977))

$$Q(\phi';\phi) = L(\phi') + H(\phi';\phi). \qquad (2bis)$$

Supposons que ϕ soit un point fixe de l'algorithme EM. D'après la relation (3.19) p. 9 de Dempster et al. (1977), le jacobien de l'opérateur T de l'algorithme EM pris au point fixe ϕ s'écrit :

$$DT(\phi) = D^{20}H(\phi;\phi)\left[D^{20}\ Q(\phi;\phi)\right]^{-1} \qquad (3)$$

où D^{20} désigne la différentielle seconde par rapport à la première variable.

D'un point de vue pratique, l'algorithme EM, facile à programmer, donne souvent de bons résultats dans des domaines très variés. Cependant, de nombreux auteurs ont souligné ses limites :

- dans certains cas, il possède de multiples points fixes qui, de plus, ne sont pas forcément des maxima locaux de la vraisemblance ; aussi, les solutions obtenues sont sensibles à sa position initiale ;

- il lui arrive assez souvent d'être d'une lenteur rédhibitoire ;

- pour certains modèles, il impose de fixer a priori des paramètres structurels importants (comme, par exemple, le nombre de composants d'un mélange de densités).

L'algorithme SEM vise à répondre à ces limitations.

2.3. L'algorithme SEM

Cet algorithme incorpore une étape stochastique (étape S) entre les étapes E et M de l'algorithme EM. Cette étape S repose sur un principe d'affectation aléatoire que nous énonçons maintenant.

Principe d'affectation aléatoire

Remplacer chaque quantité manquante par une valeur tirée au hasard suivant la densité conditionnelle $\mathbf{k}(\mathbf{y}(\mathbf{x}|\phi^n)$ *sachant* $\mathbf{x}$ *et pour un estimateur courant* ϕ^n *du paramètre.*

De cette façon, on obtient à l'étape n un échantillon $\mathbf{y}^n$ "complété", sur la base duquel on calculera un estimateur du maximum de vraisemblance de ϕ, qui fournit ϕ^{n+1}, et ainsi de suite. Partant d'une position ϕ^0, une itération de l'algorithme SEM, qui à ϕ^n associe ϕ^{n+1}, est définie ainsi :

étape E : calcul de la densité conditionnelle $\mathbf{k}(\mathbf{y}|\mathbf{x};\phi^n)$;

étape S : tirage au hasard d'un échantillon $\mathbf{z}^n$ suivant la loi de $\mathbf{k}(\mathbf{y}|\mathbf{x};\phi^n)$: on produit ainsi un échantillon complété $\mathbf{y}^n = (\mathbf{x},\mathbf{z}^n)$;

étape M : calcul de l'estimateur du maximum de vraisemblance ϕ^{n+1} sur la base de l'échantillon complété $\mathbf{y}^n$.

La suite (ϕ^n) engendrée par l'algorithme SEM ne converge pas ponctuellement, à cause des perturbations aléatoires introduites dans l'étape S. Cette suite est une chaîne de Markov homogène ; et si, pour tout ϕ et tout $\mathbf{y}$, $\mathbf{k}(\mathbf{y}|\mathbf{x};\phi)$ est strictement positive, elle est irréductible.

L'étude des propriétés de convergence de l'algorithme SEM a été menée dans le cadre de l'identification de mélanges finis de densités (cf. Celeux et Diebolt (1986), Celeux (1987) ou Diebolt (1989)). On montre que la suite (ϕ^n) est ergodique et converge vers une distribution de probabilité ψ. De plus, sous des conditions de régularité portant essentiellement sur l'opérateur T de l'algorithme EM, on montre que l'écart de ψ à l'unique solution convergente des équations de vraisemblance suit, asymptotiquement (lorsque la taille de l'échantillon $\mathbf{x}$ tend vers l'infini), une loi normale de moyenne nulle et de matrice variance non dégénérée.

D'un point de vue pratique, l'algorithme SEM répond effectivement aux limitations de l'algorithme EM (cf. Celeux et Diebolt (1986)). En particulier, ses résultats ne dépendent pas de

sa position initiale : la suite (ϕ^n) ne séjourne pas indéfiniment au voisinage d'un col de la vraisemblance et, de plus, l'algorithme SEM ne rencontre pas les situations de convergence lente de l'algorithme EM. Par ailleurs, il permet d'estimer correctement le nombre de composants d'un mélange à condition d'être initialisé avec un nombre de composants non inférieur au vrai nombre. Par contre, il est moins fiable que l'algorithme EM pour l'identification de modèles à partir de très petits échantillons ; dans ce cas, les perturbations aléatoires de l'étape S prennent trop d'importance.

Pour terminer cette section, nous allons donner une relation de récurrence qui relie de manière formelle les algorithmes EM et SEM. Cette relation nous permettra de présenter de manière synthétique l'algorithme SAEM dans la section suivante.

L'estimateur ϕ^{n+1} fourni par l'algorithme SEM à l'itération n+1 ne dépend que de ϕ^n et de $\mathbf{z}^n$. Si on note M l'opérateur qui à $\mathbf{z}^n$ associe ϕ^{n+1} via l'étape M, on peut écrire l'identité

$$\phi^{n+1} = T(\phi^n) + M(\mathbf{z}^n) - T(\phi^n)$$

où, rappelons-le, T est l'opérateur de l'algorithme EM. En posant $V(\phi^n,\mathbf{z}^n) = M(\mathbf{z}^n) - T(\phi^n)$, cette relation de récurrence s'écrit

$$\phi^{n+1} = T(\phi^n) + V(\phi^n,\mathbf{z}^n) \qquad (R)$$

3. L'algorithme SAEM

Partant d'une position ϕ^0, une itération de l'algorithme SAEM, qui à ϕ^n associe ϕ^{n+1}, est définie par l'équation

$$\phi^{n+1} = T(\phi^n) + \gamma_n\, V(\phi^n,\mathbf{z}^n) \qquad (4)$$

où la suite de réels positifs (γ_n) décroît vers 0 quand n tend vers l'infini, en partant de $\gamma_0=1$. Cette suite joue le même rôle que la température dans les algorithmes de recuit simulé. En effet, lorsque n augmente, l'importance des perturbations introduites par les $\mathbf{z}^n$ diminue.

Examinons la situation de cet algorithme vis-à-vis des algorithmes EM et SEM. Comme l'algorithme SEM, il utilise le principe d'affectation aléatoire pour produire des perturbations dans la suite (ϕ^n) des estimés du paramètre. Comme l'algorithme EM, et contrairement à l'algorithme SEM, il propose, à la convergence, un estimateur ponctuel du paramètre.

Ainsi, par rapport à l'algorithme EM, on peut penser que l'introduction de perturbations aléatoires permettra à l'algorithme SAEM d'éviter de converger vers un minimum local ou un col de la vraisemblance, et de ne pas produire de situation de convergence lente.

Par contre, par rapport à l'algorithme SEM, le fait qu'il converge ponctuellement rend son étude mathématique plus simple, et laisse espérer d'obtenir, sous des hypothèses larges, des résultats de convergence presque sûre vers un maximum local de la vraisemblance, au lieu de convergence en loi comme pour l'algorithme SEM. De plus, et pour la même raison, il s'avère plus simple à exploiter que l'algorithme SEM. Notons, cependant, que l'analyse statistique de la loi stationnaire de l'algorithme SEM fournit des indicateurs précieux pour évaluer l'incertitude attachée à l'estimation des paramètres (cf. Celeux et Diebolt (1986)). Enfin, du fait qu'il réduit l'intensité des perturbations aléatoires au fil des itérations, l'algorithme SAEM est plus fiable que l'algorithme SEM pour traiter de petits échantillons.

Dans cette section, nous n'avons pas détaillé les différentes étapes de l'algorithme SAEM. En fait, comme pour les algorithmes EM et SEM, les formules de calcul varient grandement

suivant le modèle statistique à identifier. Nous donnons une description précise de l'algorithme SAEM dans la section suivante consacrée à l'étude de son comportement pour l'identification d'un mélange fini de densités de probabilité.

4. Identification de mélanges par l'algorithme SAEM

Dans cette section, nous étudions les propriétés mathématiques de l'algorithme SAEM pour l'identification de mélanges de densités de probabilité. Ce problème est, en effet, celui pour lequel l'algorithme EM est le plus utilisé (cf. par exemple, Titterington, Smith et Makov (1985) ; c'est encore plus vrai pour l'algorithme SEM, qui a été primitivement conçu dans ce cadre (cf. Celeux et Diebolt (1986), et aussi Tassi (1989)).

4.1. Description de l'algorithme

On dispose d'un échantillon $\mathbf{x} = (x_1,...,x_N)$ d'une variable aléatoire à valeurs dans $\mathbf{R}^d$ de densité

$$h(x) = \sum_{k=1}^{K} p_k \, h(x,a_k)$$

où les proportions du mélange p_k, $k=1,...,K$ sont strictement positives et vérifient l'équation $p_1+...+p_K=1$; et où $h(.,a)$ est une densité de probabilité de forme paramétrique spécifiée, $a \in \mathbf{R}^s$ désignant le paramètre.

Il s'agit d'estimer le paramètre $\phi = (p_1,...,p_K,a_1,...,a_K)$. Avant de décrire l'algorithme SAEM, nous allons préciser la forme des entités en jeu dans le formalisme décrit en 2.1. L'échantillon complet $\mathbf{y} = (\mathbf{x},\mathbf{z})$ avec $\mathbf{z} = (z_1,...,z_N)$ où $z_i = (z_{ij},\ j=1,...,K)$ est un vecteur indicateur dont le $k^{\text{ième}}$ coordonnée vaut 1 si x_i est issu du $k^{\text{ième}}$ composant du mélange et dont les autres coordonnées sont nulles. Soit $p(z_i) = p_k$ et $a(z_i) = a_k$ si $z_i = (0,...,1,...,0)$ le 1 étant dans la $k^{\text{ième}}$ position. On a (cf., par exemple, Titterington et al. (1985))

$$g(y;\phi) = \prod_{i=1}^{N} p(z_i) \, h(x_i,a(z_i))$$

et la formule (1) s'écrit

$$f(x;\phi) = \prod_{i=1}^{N} \sum_{k=1}^{K} p_k \, h(x_i,a_k)$$

De plus, $k(y|x;\phi) = \prod_{i=1}^{N} k(y_i|x_i;\phi)$ avec $y_i = (x_i,z_i)$ et $k(y_i|x_i;\phi) = \dfrac{p(z_i)\, h(x_i,a(z_i))}{\sum_{\ell=1}^{K} p_\ell \, h(x_i,a_\ell)}$,

et si nous notons $t_k(x_i) = p_k \, h(x_i,a_k) \,/\, \sum_{\ell=1}^{K} p_\ell \, h(x_i,a_\ell)$ la probabilité a posteriori qu'un point x_i soit issu du $k^{\text{ième}}$ composant, on a

$$Q(\phi';\phi) = \sum_{i=1}^{N} \sum_{\ell=1}^{K} t_\ell(x_i) \left\{ \ell n \; p'_\ell + \ell n \; h(x_i,a'_\ell) \right\} .$$

Indiquons, enfin, que dans la suite nous noterons $e(x_i) = z_i$ afin de faire ressortir le lien entre une observation et sa provenance et de respecter des notations bien établies pour l'algorithme SEM dans le cadre des mélanges. Une itération de l'algorithme SAEM, qui à ϕ^n associe ϕ^{n+1}, se décompose ainsi.

Etape E : calcul des probabilités a posteriori $t_k^n(x_i)$, k=1,...,K, sachant ϕ^n, d'appartenance des points (x_i, i=1,...,N) de l'échantillon à chaque composant du mélange.

Etape S : tirage pour chaque x_i de la variable $e^n(x_i) = (e_k^n(x_i),\ k=1,...,K)$ selon une loi multinomiale d'ordre un et de paramètre $(t_k^n(x_i),\ k=1,...,K)$. S'il existe k tel que card $\{x_i/e_k^n(x_i) = 1\}$ soit plus petit que (d+1)/N, l'algorithme est réinitialisé au hasard sur la base de K-1 composants. Sinon

Etape A : calcul des quantités $r_k^n(x_i) = t_k^n(x_i) + \gamma_n(e_k^n(x_i) - t_k^n(x_i))$ pour i=1,...,N ; k=1,...,K

Etape M : calcul de ϕ^{n+1}, estimateur du maximum de vraisemblance de ϕ, les $r_k^n(x_i)$ étant considérés artificiellement comme les probabilités a posteriori d'appartenance des points x_i aux composants du mélange. Pour k=1,...,K on en déduit $p_k^{n+1} = \Sigma\ r_k^n(x_i)/N$ et cela conduit à résoudre les équations de vraisemblance $\Sigma\, r_k^n\, \{\partial\, \ell n\, h(x_i,a_k)/\partial a_k\} = 0$.

Le comportement théorique et pratique de l'algorithme SAEM dépend bien sûr fortement de la vitesse de convergence vers 0 et de la décroissance régulière de la suite (γ_n). Cette question est discutée dans la suite.

4.2. Convergence de l'algorithme SAEM

La suite de la section est consacrée à l'étude de la convergence de l'algorithme SAEM pour des mélanges dont les composants ont des densités appartenant à la famille exponentielle. Avant d'établir un théorème décrivant la convergence presque sûre de cet algorithme vers l'un des points fixes attractifs de l'algorithme EM, nous donnons quelques résultats préliminaires.

4.2.1. Résultats préliminaires

Le mécanisme de réinitialisation, décrit à l'étape S de l'algorithme, assure que la suite (ϕ^n) reste dans un compact G sur lequel la vraisemblance L est de classe C^2 et est donc bornée.

Le fait de se restreindre aux mélanges de la famille exponentielle facilite l'étude mathématique de l'algorithme et ne limite pas la portée réelle des résultats obtenus car tous les mélanges considérés en pratique sont constitués de densités de ce type. Rappelons que la famille paramétrée de densités h(x,a) appartient à la famille exponentielle si

$$h(x,a) = d(a)^{-1}\ n(x)\ \exp({}^t a\ b(x)) \qquad x \in \mathbf{R}^d$$

où $n : \mathbf{R}^d \to \mathbf{R}$, $b : \mathbf{R}^d \to \mathbf{R}^s$ et où $d(a) = \int_{\mathbf{R}^d} n(x)\ \exp({}^t a\ b(x))\ dx$.

Dans ce cas, une itération de l'algorithme EM se résume par les formules (cf. Redner et Walker 1984), pour k=1,...,K

$$p_k^{n+1} = \sum_{i=1}^{N} t_k^n(x_i)/N \quad \text{et} \quad a_k^{n+1} = \sum_{i=1}^{N} t_k^n(x_i)\, b_k(x_i) / \sum_{i=1}^{N} t_k^n(x_i).$$

L'opérateur T de l'algorithme EM est continûment différentiable sur G et les points fixes de T sont les points stationnaires de la vraisemblance L (cf. Redner et Walker 1984). Ainsi, si ϕ^* est un point fixe de T, on a pour tout $\phi \in G$, proche de ϕ^*, le développement

$$T(\phi) - \phi^* = DT(\phi^*)\,(\phi - \phi^*) + O(\|\phi - \phi^*\|^2) \qquad (5)$$

$\|.\|$ étant une norme quelconque sur l'espace où les paramètres prennent leurs valeurs.

Maintenant, on a la proposition suivante :

Proposition :

(i) Pour tout point fixe ϕ^* de T, $DT(\phi^*)$ est diagonalisable et ses valeurs propres sont positives.

(ii) Tout point fixe répulsif de T est un minimum local strict de la vraisemblance, tout point fixe hyperbolique de T est un col de la vraisemblance et tout point fixe attractif de T est un maximum local strict de la vraisemblance.

Démonstration :

(i) D'après (3), $DT(\phi^*) = D^{20}H(\phi^*;\phi^*)[D^{20}Q(\phi^*;\phi^*)]^{-1}$; nous allons montrer que la matrice $D^{20}H(\phi^*,\phi^*)$, qui est symétrique, est négative et que $D^{20}(\phi^*;\phi^*)$, également symétrique, est définie négative. On pourra écrire $[D^{20}(\phi^*;\phi^*]^{-1}$ comme le produit d'une matrice régulière C et de sa transposée ${}^tC$. On montre alors facilement que $DT(\phi^*)$ a les mêmes valeurs propres que la matrice définie positive ${}^tC\, D^{20}H(\phi^*;\phi^*)C$, ce qui démontre (i). De l'expression de $Q(\phi';\phi)$ donnée en 4.1, on tire que $D^{20}Q(\phi^*;\phi^*)$ est une matrice dont les blocs non diagonaux sont nuls et dont les blocs diagonaux sont donnés par

$$\frac{\partial^2 Q(\phi^*;\phi^*)}{\partial p_\ell^{*2}} = - \sum_{i=1}^{N} t_\ell^*(x_i) / p_\ell^{*2} ;$$

$$\frac{\partial^2 Q(\phi^*;\phi^*)}{\partial a_\ell^{*2}} = \sum_{i=1}^{N} t_\ell^*(x_i)\, \frac{\partial^2 \ell n\, h(x_i;a_\ell^*)}{\partial a_\ell^{*2}}$$

avec $\quad t_\ell^*(x_i) = p_\ell^*\, h(x_i,a_\ell^*) \,/\, \sum_{k=1}^{K} p_k^*\, h(x_i,a_k^*)$

Du fait que les $h(x_i,a_\ell)$ appartiennent à la famille exponentielle, $\partial^2 \ell n\, h(x_i,a_\ell^*)/\partial a_\ell^{*2}$ est l'opposée d'une matrice variance non dégénérée.

Finalement, $\partial^2 Q(\phi^*;\phi^*)/\partial \phi^{*2}$ est une matrice diagonale par blocs tous définis négatifs, et est donc définie négative. Quant à $D^{20}H(\phi^*;\phi^*)$, du fait de la relation (2bis), et par un calcul calqué sur celui de Louis (1982 Appendix p. 232), il peut s'écrire

$$D^{20}H(\phi^*;\phi^*) = \sum_{i=1}^{N} \left\{ S^*(x_i,\phi^*)\ {}^tS^*(x_i,\phi^*) - \sum_{\ell=1}^{K} t_\ell^*(x_i)\ S(x_i,\phi_\ell^*)\ {}^tS(x_i,\phi_\ell^*) \right\}$$

où

$$S(x_i,\phi_\ell^*) = \frac{\partial\, \ell n\, p_\ell^*\, h(x_i,a_\ell^*)}{\partial \phi_\ell^*} \; ; \; S^*(x_i,\phi^*) = \sum_{\ell=1}^{K} t_\ell(x_i)\, \frac{\partial\, \ell n\, p_\ell^*\, h(x_i,a_\ell^*)}{\partial \phi_\ell^*} .$$

Par l'inégalité de Cauchy-Schwarz, la matrice $D^{20}H(\phi^*;\phi^*)$ est donc bien négative.

(ii) Ce point est une conséquence directe de (i) et de la relation (cf. Dempster et al. (1977 p. 10))

$$D^2 L(\phi^*) = \{I - DT(\phi^*)\}\, D^{20}Q(\phi^*;\phi^*).$$

4.2.2. Un théorème de convergence

Le théorème que nous allons maintenant établir concerne un modèle simplifié de l'algorithme SAEM par rapport à son fonctionnement réel tel que l'expérience et le calcul théorique le révèlent. Mais ce modèle constitue une bonne représentation de l'algorithme SAEM. Les hypothèses de ce modèle sont les suivantes :

(H1) les points fixes de T dans G sont des points isolés,

(H2) tout point fixe attractif de T appartient à G,

(H3) il existe $\rho>0$ tel que, pour tout ϕ de G, pout tout vecteur u de norme un et pour tout r satisfaisant $0 < r \leq \rho$, on ait $T(\phi) + ru \in G$,

(H4) pour tout $\phi \in G$, l'ensemble des $T(\phi) + V(\phi,e)$, quand e décrit l'ensemble des tirages aléatoires possibles, est situé de part et d'autre de tout hyperplan contenant $T(\phi)$.

L'hypothèse (H3) entraîne que dès que n dépasse une certaine valeur n_0, $\phi^{n+1}=T(\phi^n) + \gamma_n V(\phi^n,e^n)$ appartient à G pour presque tout tirage $e^n = (e^n(x_1),...,e^n(x_N))$.

Les hypothèses (H2) et (H3) traduisent le fait que le nombre de composants du mélange est connu et que la taille de l'échantillon est suffisante pour éviter l'apparition de points fixes de T au bord de G.

L'hypothèse (H4) est reliée à (H3). En effet, (H3) assure que, pour tout ϕ, $T(\phi)$ n'est pas trop près du bord de G ; et par suite, $T(\phi)$ est environné de points de G de la forme $T(\phi) + V(\phi,e)$, c'est-à-dire de points de G accessibles par l'algorithme SEM. Cette hypothèse, de formulation technique, n'est pas contraignante : le nombre de points accessibles par l'algorithme SEM est de l'ordre de K^N (K étant le nombre de composants et N la taille de l'échantillon), il est donc très grand dès que N permet une estimation réaliste des composants du mélange.

L'hypothèse (H1) ne prend pas en compte le cas où l'ensemble des points fixes de T contient une composante connexe non réduite à un point. Or, on sait que cette situation se produit (cf. Celeux (1987 p. 56)). Notons, toutefois, qu'une telle composante connexe

possède, sans doute, au moins une direction répulsive puisqu'elle conduit à des composants du mélange confondus, les proportions étant quelconques.

Théorème : Sous les hypothèses (H1) à (H4) et si la suite (γ_n) est de la forme c/n^u pour $n > 0$, avec $0 < \mu < 1$, alors la suite (ϕ_n) engendrée par l'algorithme SAEM converge p.s. vers un maximum local de la vraisemblance.

Le point fondamental dans l'hypothèse énoncée sur la suite (γ_n) est que la série $\Sigma\gamma_n$ diverge. Le fait que $\gamma_n = c/n^\mu$ est une hypothèse purement technique qui pourrait être omise au prix d'un alourdissement notable de la démonstration du théorème.

Ce théorème montre bien l'intérêt de l'algorithme SAEM par rapport à l'algorithme EM. En effet, celui-ci peut converger vers un col de la vraisemblance. D'autre part, toujours dans le cadre des mélanges de la famille exponentielle, Redner et Walker (1984) ont montré que, sous des conditions assez larges, lorsque la taille N de l'échantillon tendait vers l'infini, l'algorithme EM convergeait vers l'unique solution convergente des équations de vraisemblance, à condition que sa position initiale soit suffisamment proche de cette solution optimale. Notre théorème autorise à penser que, lorsque N tend vers l'infini, l'algorithme SAEM augmente les chances d'atteindre effectivement cette estimation optimale, quelle que soit sa position initiale.

Démonstration :

Choisissons une norme sur l'espace des paramètres ϕ. Notons F l'ensemble des points fixes de T ; à tout ϕ de F et à tout entier n on associe la boule ouverte $V_n(\phi)$ de centre ϕ et de rayon $C_v\,\gamma_n^{1/2}$, C_v étant une constante que l'on déterminera au cours de la démonstration.

Notons $A_n = G - \bigcup_{\phi \in F} V_n(\phi)$; A_n fermé de G est compact. Nous allons montrer qu'après chaque séjour dans A_n, la suite (ϕ_n) quitte A_n en un temps fini p.s. Puis, nous montrerons que la suite (ϕ_n) étant, éventuellement, entrée dans le voisinage d'un point fixe possédant au moins une direction répulsive, elle en sortira définitivement en un temps fini. Il en résultera que (ϕ_n) finira par rester définitivement dans le voisinage $V_n(\phi)$ d'un point fixe attractif ϕ ; ce qui terminera la démonstration d'après la proposition de 4.2.1.

Etape 1 : ϕ_n sort de A_n en un temps fini. La fonction $(LoT - L)(\phi)$ définie sur G à valeurs dans $\mathbf{R}^+$ est de classe C^2 sur un ouvert contenant G et ne s'annule que si $\phi \in F$: les éléments de F sont les minima absolus de cette fonction et, donc, si $\phi \in F$, $D(LoT - L)(\phi) = 0$ et $D^2(LoT - L)(\phi)$ est définie positive.

Comme F est fini, on en déduit qu'il existe une constante $\alpha > 0$ telle que

$$(\forall h) \qquad D^2(LoT - L)(\phi)(h,h) \geq 4\alpha \, ||h||^2 \tag{I1}$$

En effet, soit λ_{min} la plus petite valeur propre de $D^2(LoT - L)(\phi)$

$$(\forall h) \qquad D^2(LoT - L)(\phi)(h,h) \geq \lambda_{min} \, ||h||_I^2$$

où $\|.\|_I$ désigne la norme euclidienne usuelle ; les normes étant équivalentes, il existe donc $\alpha(\phi) > 0$ tel que $D^2(LoT - L)(\phi)(h,h) \geq 4\alpha(\phi)\|h\|^2$. F étant fini, il suffit de prendre $\alpha = \min_{\phi \in F} \alpha(\phi)$ pour obtenir (I1).

De (I1) on déduit que, pour n assez grand,

$$(\forall\ \phi \in F) \qquad (\forall\ h \text{ avec } \|h\| = C_v \gamma_n^{1/2}) \qquad (LoT - L)(\phi+h) \geq \alpha\, C_v^2\, \gamma_n \qquad (I2)$$

En effet, dès que $\|h\| = C_v \gamma_n^{1/2} < \varepsilon(\phi)$, avec $\varepsilon(\phi)$ petit, le développement de $(LoT - L)(\phi+h)$ autour de ϕ donne

$$(LoT - L)(\phi+h) = (1/2)\, D^2(LoT - L)(\phi)(h,h) + O\|h\|^3$$

ainsi

$$(LoT - L)(\phi+h) \geq 2\alpha\, \|h\|^2 - \text{Cste}\, \|h\|^3$$

où la notation Cste désigne une constante strictement positive.

Si $\varepsilon(\phi)$ est choisi assez petit, on aura $(LoT - L)(\phi+h) \geq \alpha\|h\|^2$. Il suffit donc de prendre n assez grand pour que $C_v \gamma_n^{1/2} < \min_{\phi \in F} \varepsilon(\phi)$ pour obtenir (I2).

Maintenant, LoT - L est strictement positive sur le compact A_n, et atteint sur A_n sa borne inférieure $c_n > 0$.

Nous allons montrer que, pour n assez grand, c_n est plus grand que $\alpha\, C_v^2\, \gamma_n$. Démontrons-le par l'absurde. Si ce n'était pas vrai, il existerait une suite (g_n) de G telle que l'on ait, à partir d'un certain rang, $g_n \in A_n$ et $\varepsilon_n = (LoT-L)(g_n) < \alpha\, C_v^2\, \gamma_n$. Par compacité de G, (g_n) aurait une valeur d'adhérence g dans G et on pourrait extraire une sous-suite $g_{\varphi(n)}$ qui convergerait vers g. Par continuité de LoT - L, $(LoT - L)(g_{\varphi(n)})$ convergerait vers $(LoT - L)(g) = 0$ et donc g appartiendrait à F. Il en résulterait que, d'après (I2), on aurait $\varepsilon_{\varphi(n)} \geq \alpha\|g - g_{\varphi(n)}\|^2$. Comme par hypothèse $\varepsilon_{\varphi(n)} < \alpha C_v^2\, \gamma_{\varphi(n)}$, on aurait ainsi $\|g-g_{\varphi(n)}\| < C_v\, \gamma_{\varphi(n)}$ et donc $g_{\varphi(n)} \in V_{\varphi(n)}(g)$ ce qui contredirait le fait que $g_{\varphi(n)} \in A_{\varphi(n)}$.

Enfin, par l'égalité des accroissements finis, on peut écrire

$$L(\phi^{n+1}) - L(\phi^n) = (LoT - L)(\phi^n) + \gamma_n\, DL(\phi^n + t\gamma_n\, V(\phi^n,e^n))\, V(\phi^n,e^n)$$

pour un élément t de [0,1].

Posons $SUP = \sup_{g \in G, e} \|V(g,e)\|$; si $\phi^n \in A_n$ on a, pour n assez grand,

$$L(\phi^{n+1}) - L(\phi^n) \geq \alpha\, C_v^2\, \gamma_n - \gamma_n \sup_{g \in G} \|DL(g)\|\, SUP$$

et si l'on choisit C_v assez grand pour que la quantité $\alpha C_v^2 - \sup_{g\in G} \|DL(g)\|$ SUP soit strictement positive, on obtient

$$L(\phi^{n+1}) - L(\phi^n) \geq \text{Cste } \gamma_n .$$

Comme $\Sigma \gamma_n = \infty$, ϕ^n ne peut rester indéfiniment dans A_n, sinon $L(\phi^{n+1}) - L(\phi^0)$ tendrait vers l'infini, or L est bornée sur le compact G.

Etape 2 : ϕ_n ne peut rester indéfiniment dans le voisinage $V_n(\phi)$ d'un point fixe ϕ possédant au moins une direction répulsive.

Cette étape va elle-même se décomposer en plusieurs points.

Point 1: on va d'abord montrer que si ϕ^n est entré dans le voisinage $V_n(\phi^*)$ d'un point fixe ϕ^* possédant une direction répulsive, alors presque sûrement, ϕ^n sortira en un temps fini de la boule B_n de centre ϕ^* et de rayon $K\gamma_n$, K étant une constante positive à préciser.

Point 2: on montre ensuite que l'on peut choisir la constante K ci-dessus pour que, dès que ϕ^n a quitté la boule B_n, l'on soit assuré que, presque sûrement, au bout d'un temps fini, ϕ^{n+k} aura quitté la boule $HV_{n+k}(\phi^*)$ de centre ϕ^* et de rayon $HC_v \gamma_{n+k}^{1/2}$, H étant une constante plus grande que 1 à préciser.

Point 3: on montre que l'on peut choisir la constante H ci-dessus pour que, dès que ϕ^n a quitté $HV_n(\phi^*)$, l'on soit assuré que la suite (ϕ^m) ne rentrera plus jamais dans le voisinage $V_m(\phi^*)$. Ce qui termine la démonstration.

Démonstration du point 1.

Soit ϕ^* un point fixe de T possédant au moins une direction répulsive ; soit v un vecteur propre de norme 1 associé à une valeur propre $\lambda > 1$ de l'opérateur linéaire $DT(\phi^*)$. On suppose que ϕ_n appartient à la boule B_n de centre ϕ^* et de rayon $K\gamma_n$; il s'agit de montrer, qu'au bout d'un temps fini j, la suite (ϕ_m) sort de cette boule. On note

$$q_n = \langle v, \phi^n - \phi^* \rangle \text{ et } w_n = \langle v, V(\phi^n, e^n) \rangle$$

où <.,.> désigne le produit scalaire usuel.

On considère des évènements $E_{n,j} = E^+_{n,j} \cup E^-_{n,j}$, $j \geq 1$, où les évènements $E^+_{n,j}$ et $E^-_{n,j}$ sont de la forme

$$E^+_{n,j} = \{q_n \geq 0,\ w_n > a, \ldots, w_{n+j} > a\}$$

$$E^-_{n,j} = \{q_n \leq 0,\ w_n < -a, \ldots, w_{n+j} < -a\}$$

a étant une constante strictement positive à préciser.

On montre en annexe que, pour tout entier j fixé, il existe une valeur de a > 0 pour laquelle l'évènement $E_{n,j}$ ait lieu infiniment souvent p.s. Supposons maintenant que ϕ^n appartienne à la boule B_n ; d'après la relation (5), il vient

$$|q_{n+1}| \geq \lambda\,|q_n| + O\|\phi^n - \phi^*\|^2 + \gamma_n\|w_n\|$$
$$|q_{n+1}| \geq \gamma_n\|w_n\| - C\gamma_n^2,\ C \text{ étant une constante positive.}$$

On peut se restreindre au cas où l'entier n est assez grand pour garantir que le facteur $(1 - c\gamma_n/a)$ soit plus grand que 1/2. Si, dans ces conditions, ϕ^n appartient à B_n, alors, dès qu'un évènement $E_{m,j}$, $m \geq n$, se produit, si ϕ^m n'est pas déjà sorti de B_m, on a $|q_{m+1}| \geq a\,\gamma_m(1 - c\gamma_m/a)$; on aura donc $|q_{m+j}| \geq K\,\gamma_{m+j}$ pourvu que l'on ait pris soin d'avoir préalablement déterminé l'entier j de telle manière que $\lambda^{j-1} \geq 2K/a$. Pour assurer cette inégalité, il suffit d'avoir choisi $(j - 1) \geq \ell n\,(2K/a)/\ell n\,\lambda$.

Démonstration du point 2.

Il s'agit de montrer que l'on peut choisir K pour assurer qu'en un temps fini la suite (ϕ^n) sort de la boule $HV_n(\phi^*)$ de centre ϕ^* et de rayon $HC_v\,\gamma_n^{1/2}$, $H > 1$.

Si $\|\phi^n - \phi^*\| < HC_v\,\gamma_n^{1/2}$, on a $|q_n| < HC_v\,\gamma_n^{1/2}$. Tant que $K\gamma_n < |q_n| < HC_v\,\gamma_n^{1/2}$, on a

$$|q_{n+1}| \geq \lambda K\gamma_n - H^2C_v^2\,\gamma_n - SUP\,\gamma_n$$
$$|q_{n+1}| \geq (\lambda K - H^2C_v^2 - SUP)\,\gamma_n$$

Si K a été choisi assez grand pour que $\lambda K - H^2C_v^2 - SUP \geq (1 + \varepsilon)K$, ε étant une constante positive telle que $\lambda > 1 + \varepsilon$, c'est-à-dire que l'on a choisi

$$K \geq \frac{H^2C_v^2 + SUP}{\lambda(1+\varepsilon)}$$

on a

$$|q_{n+1}| \geq (1 + \varepsilon)\,K\gamma_{n+1},\ \ldots,\ |q_{n+k}| \geq (1 + \varepsilon)^k\,K\gamma_{n+k}.$$

Donc, dès que k dépasse une valeur suffisante, on est assuré que $|q_{n+k}| > HC_v\,\gamma_{n+k}^{1/2}$.

Démonstration du point 3.

Rappelons que $V_n(\phi^*)$ désigne la boule de centre ϕ^* et de rayon $C_v\gamma_n^{1/2}$ et que l'on note $HV_n(\phi^*)$ la boule de centre ϕ^* et de rayon $HC_v\gamma_n^{1/2}$ $(H > 1)$.

Nous allons montrer que l'on peut choisir H de telle sorte que, au premier instant n^* de sortie de $HV_{n^*}(\phi^*)$, avec n^* assez grand,

$$L(\phi^{n^*}) > \sup_{\phi \in V_{n^*(\phi^*)}} L(\phi)$$

et, qu'à partir de là, $L(\phi^n)$ croît strictement tant que ϕ^n reste proche de ϕ^* ; et donc, que la suite (ϕ^n) ne peut pas revenir dans $V_{n^*}(\phi^*)$.

Ainsi, par un raisonnement analogue à celui de l'étape 1, la suite (ϕ^n) entrera dans le voisinage de $V_n(\phi)$ d'un point fixe attractif ϕ et, vu que $V(\phi, e)$ est uniformément borné, si n est devenu assez grand, les ϕ^{n+m} resteront dans les voisinages $V_{n+m}(\phi)$ indéfiniment.

Etablissons, tout d'abord, quelques inégalités.
Pour n assez grand, si $g \notin V_n(\phi^*)$ mais $\|g - \phi^*\| < \varepsilon$, avec ε assez petit, on a

$$L \circ T(g) - L(g) \geq \alpha C_v^2 \gamma_n. \tag{I4}$$

De même, pour n assez grand, si $g \notin HV_n(\phi^*)$ mais $\|g - \phi^*\| < \varepsilon$, on a

$$L \circ T(g) - L(g) \geq \alpha H^2 C_v^2 \gamma_n. \tag{I5}$$

Ces deux inégalités se démontrent de la même manière que (I2). Maintenant, pour n assez grand,

$$\sup_{g \in V_{n^*(\phi^*)}} |L(g) - L(\phi^*)| \leq C_{DL} C_v^2 \gamma_n. \tag{I6}$$

où $C_{DL} = \sup_{g \in G} \|D^2 L(g)\|$.

Cette inégalité vient de ce que $DL(\phi^*) = 0$ et donc que

$$|L(g) - L(\phi^*)| \leq \frac{1}{2} \|D^2 L(\phi^*)\| \, \|g - \phi^*\|^2 + O\|g - \phi^*\|^3.$$

On montre de même que

$$\sup_{g \in V_{n^*(\phi^*)}} |L \circ T(g) - L \circ T(\phi^*)| \leq C_{DLT} C_v^2 \gamma_{n^*}. \tag{I'6}$$

où $C_{DLT} = \sup_{g \in G} \|D^2 (L \circ T)(g)\|$

Enfin, on peut choisir la constante C_v assez grande pour que l'on ait, pour n assez grand,

$$\text{dès que } \phi^n \notin V_n(\phi^*),\ L(\phi^{n+1}) - L(\phi^n) > \frac{\alpha}{2} C_v^2 \gamma_n. \tag{I7}$$

En effet, on a
$L(\phi^{n+1}) - L(\phi^n) = L(\phi^{n+1}) - L(T(\phi^n)) + L(T(\phi^n)) - L(\phi^n)$

$$= (L \circ T - L)(\phi^n) + L\{T(\phi^n) + \gamma_n V(\phi^n, e^n)\} - L(T(\phi^n))$$
$$= (L \circ T - L)(\phi^n) + \gamma_n DL\{T\phi^n) + t\,\gamma_n V(\phi^n, e^n)\}\, V(\phi^n, e^n) \quad t \in [0, 1]$$
(par l'égalité des accroissements finis)
$$\geq \alpha C_v^2 \gamma_n - \sup_{g \in G} \|DL(g)\|\, SUP\, \gamma_n$$
(par I4)

il reste donc à choisir C_v assez grand pour que

$$\alpha C_v^2 - \sup_{g \in G} \|DL(g)\|\, SUP \geq \frac{\alpha}{2} C_v^2.$$

L'inégalité (I7), vraie également (mutatis mutandis) dès que $\phi^n \notin HV_n(\phi^*)$, nous assure que la vraisemblance $L(\phi)$ croît strictement lorsque ϕ^n est hors de $HV_n(\phi^*)$, avec $\|\phi^n - \phi^*\| < \varepsilon$.

Maintenant, il existe un instant k (nous serons amenés à supposer que k est assez grand) auquel ϕ^k sort de $V_k(\phi^*)$ et ϕ^ℓ ne rentre plus dans $V_\ell(\phi^*)$ pour $k < \ell < n^*$. Rappelons que n^* est l'instant de sortie de $HV_{n^*}(\phi^*)$ que l'on considère. Il sera commode de noter $n^* = k + p$. Pour minorer $L(\phi^{n*})$, nous allons utiliser l'identité

$$L(\phi^{k+p}) - L(\phi^*) = \{L(\phi^{k+p}) - L(\phi^{k+p-1})\} + \{L(\phi^{k+p-1}) - L(\phi^k)\} + \{L(\phi^k) - L(\phi^*)\} \quad (E)$$

Partant de cette identité, nous allons minorer chaque terme entre accolades.

1) Minoration de $L(\phi^k) - L(\phi^*)$: on a $L(\phi^k) - L(\phi^*) = L(\phi^k) - L(\phi^{k-1}) + L(\phi^{k-1}) - L(\phi^*)$; comme $\phi^{k-1} \in V_{k-1}(\phi^*)$ et par application de (I6), on a
$L(\phi^k) - L(\phi^*) \geq - C_{DL} C_v^2 \gamma_{k-1} - |L(\phi^*) - L(\phi^{k-1})|$

et $|L(\phi^k) - L(\phi^{k-1})| \leq |L(\phi^k) - L(T(\phi^{k-1}))| + |L(T(\phi^{k-1})| - L(\phi^{k-1})|$
$\leq \|DL(T(\phi^{k-1}))\, \gamma_{k-1}\, V(\phi^{k-1}, e^{k-1})\|$
$+ \frac{1}{2} \gamma_{k-1}^2 \sup_{g \in G} \|D^2 L(g)\|\, (SUP)^2 + \sup_{g \in V_{k-1}(\phi^*)} |L(T(\phi)) - L(\phi)|$

Or T est lipschitzienne car DT est continue, et donc bornée, sur le compact G ; soit C_T la constante de Lipschitz correspondante. De même, DL est lipschitzienne car $D^2 L$ est continue sur G, et la constante de Liptschitz est C_{DL}. Ainsi,

$$\|DL(T(\phi^{k-1})) - DL(T(\phi^*))\| \leq C_{DL}\, C_T\, C_v\, \gamma_{k-1}^{1/2}, \text{ avec } DL\,(T(\phi^*) = 0.$$

Donc

$$\|DL(T(\phi^{k-1}))\, \gamma_{k-1}\, V(\phi^{k-1}, e^{k-1})\| \leq C_{DL}\, C_T\, C_v\, SUP\, \gamma_{k-1}^{3/2}.$$

Par ailleurs, pour k assez grand, par (I6) et (I'6), on a

$$\sup_{g \in V_{k-1}(\phi*)} |L(T(g)) - L(g)| \leq (C_{DL} + C_{DLT})\, C_v^2\, \gamma_{k-1}.$$

Finalement, pour k assez grand

$$|L(\phi^k) - L(\phi^{k-1})| \leq C_{DT}\, C_T\, C_V\, SUP\, \gamma_{k-1}^{3/2} + \frac{1}{2}\gamma_{k-1}^2 \sup_{g\in G} \|D^2 L(g)\|\, (SUP)^2$$
$$+ (C_{DL} + C_{DLT})\, C_v^2\, \gamma_{k-1}$$

Mais, pour k assez grand, $\gamma_{k-1} \leq 3/2\ \gamma_k$, tandis que $\gamma_{k-1}^{3/2} = o(\gamma_{k-1})$ et $\gamma_{k-1}^2 = o(\gamma_{k-1})$; donc

$$|L(\phi^k) - L(\phi^{k-1})| \leq 2(C_{DL} + C_{DLT})\, C_v^2\, \gamma_k.$$

Finalement, comme $\gamma_{k-1} \leq 2\gamma_k$,

$$L(\phi^k) - L(\phi^*) \geq -(4C_{DL} + 2C_{DLT})\, C_v^2\, \gamma_k. \qquad (I8)$$

2) Considérons maintenant $L(\phi^{k+p}) - L(\phi^k)$. D'après (I7)

$$L(\phi^{k+p}) - L(\phi^k) = \sum_{m=1}^{k+p-1} L(\phi^{m+1}) - L(\phi^m) \geq (\alpha/2)\, C_v^2 \sum_{m=1}^{k+p-1} \gamma_m.$$

Nous allons utiliser ici le fait que $\gamma_m = c/m^\mu$, $0 < \mu < 1$, pour $m \geq 1$, en posant $c = 1$ pour simplifier. On a l'équivalence classique suivante, lorsque k tend vers l'infini et que l'entier p reste plus petit qu'une borne fixe p_0 que nous préciserons plus loin

$$\sum_{m=1}^{k+p-1} \gamma_m \sim \int_k^{k+p} 1/t^\mu\, dt = \frac{1}{1-\mu}\{(k+p)^{1-\mu} - k^{1-\mu}\}.$$

Ainsi $L(\phi^{k+p}) - L(\phi^k)$ est-il minoré par une quantité équivalente, lorsque k tend vers l'infini, à $(\alpha/2)\, C_v^2\, k^{1-\mu}\ \{(1+p/k)^{1-\mu} - 1\}$. Mais, lorsque k tend vers l'infini et que $0 \leq p \leq p_0$, on a $(1+p/k)^{1-\mu} = 1 + (1-\mu)\, p/k + o(1/k^2)$ et cette dernière quantité reste plus grande que $(1-\mu)\, p/(2k)$ pour k est assez grand.

Dans ces conditions, $L(\phi^{k+p}) - L(\phi^k)$ est minoré par une quantité équivalente, lorsque k tend vers l'infini et que $0 \leq p \leq p_0$, à une quantité minorée par

$$(\alpha/2)\, C_v^2\, k^{1-\mu}\, (1-\mu)\, p/(2k) = \frac{\alpha(1-\mu)\, C_v^2}{4}\, p\, \gamma_k.$$

Donc, quitte à choisir k assez grand, on obtient l'inégalité, vraie pour tout p, $0 \leq p \leq p_0$:

$$L(\phi^{k+p}) - L(\phi^k) \geq \frac{\alpha(1-\mu)\, C_v^2}{8}\, p\, \gamma_k. \qquad (I9)$$

Remarquons maintenant que, puisque $L(\phi^{\ell+1}) > L(\phi^\ell)$ pour tout ℓ compris entre k et n^*, on peut déduire de (I9) que si $p > p_0$ on a cette fois pour k assez grand :

$$L(\phi^{k+p}) - L(\phi^k) \geq L(\phi^{k+p_0}) - L(\phi^k) \geq \frac{\alpha(1-\mu)\, C_v^2}{8}\, p_0\, \gamma_k. \qquad (I9bis)$$

A présent, nous allons précisément chosir p_0 assez grand pour assurer que, chaque fois que $p > p_0$, et pour des valeurs de k assez grandes, la quantité $L(\phi^{k+p}) - L(\phi^*)$ dépasse

$$2 \sup_{g \in V_{k+p}(\phi^*)} |L(g) - L(\phi^*)| + (4C_{DL} + 2C_{DLT})\, C_v^2\, \gamma_k.$$

Pour cela, d'après (I9bis), (I8) et (I6), il suffit de choisir p_0 de manière que l'on ait (par exemple)

$$\frac{\alpha(1-\mu)\, C_v^2}{8}\, p_0\, \gamma_k \geq 2(4C_{DL} + 2C_{DLT})\, C_v^2\, \gamma_k$$

$$\text{i.e.} \quad p_0 \geq 16(4C_{DL} + 2C_{DLT}) / \{\alpha(1-\mu)\} \qquad (I10)$$

Maintenant, de deux choses l'une :

- ou bien $p > p_0$, et alors le choix de p_0 ci-dessus permet de conclure :

$$L(\phi^{k+p}) > L(\phi^*) + 2 \sup_{g \in V_{k+p}(\phi^*)} |L(g) - L(\phi^*)| \quad \text{CQFD ;}$$

- ou bien $1 \leq p \leq p_0$; dans ce cas, pour tout $\eta > 0$ arbitrairement petit, le rapport $(\gamma_{k+p} / \gamma_k)$ reste compris entre $1-\eta$ et 1. Pour parvenir à conclure, nous allons établir qu'alors la différence $L(\phi^{k+p}) - L(\phi^{k+p-1})$ est minoré par une quantité dont l'ordre de grandeurs est $H^2\, C_v^2\, \gamma_{k+p}$, soit $H^2\, C_v^2\, \gamma_k(1-\eta)$. On a en effet

$$L(\phi^{k+p}) - L(\phi^{k+p-1}) = (LoT - L)(\phi^{k+p-1}) + [L\{T(\phi^{k+p-1}) + \gamma_{k+p-1}\, V(\phi^{k+p-1}, e^{k+p-1})\} - L(T(\phi^{k+p-1}))].$$

La valeur absolue du terme entre crochets est, comme dans la preuve de (I8), majorée par $O(\gamma_{k+p-1}^{3/2})$ puisque $\phi^{k+p-1} \in HV_{k+p-1}(\phi^*)$. Quant au terme $(LoT - L)(\phi^{k+p-1})$, il faut pour le minorer tenir compte du fait que $\phi^{k+p} \notin HV_{k+p}(\phi^*)$ et donc que $\|\phi^{k+p} - \phi^*\| > HC_v\gamma_{k+p}^{1/2}$; k

étant grand, γ_{k+p-1}, γ_{p+k} sont presque égaux et il s'ensuit que $\|T(\phi^{k+p-1}) - \phi^*\| > HC_v \gamma_{k+p}^{1/2} - o(\gamma_{k+p-1}^{1/2})$, d'où : $\|T(\phi^{k+p-1}) - \phi^*\| > (HC_v / 2)\gamma_{k+p}^{1/2}$. Or, d'autre part, on a

$$\|T(\phi^{k+p-1}) - \phi^*\| \leq C_T \|\phi^{k+p-1} - \phi^*\|$$

d'où il résulte que

$$\|\phi^{k+p-1} - \phi^*\| > \frac{HC_v}{2C_T}\gamma_{k+p}^{1/2} \qquad (I11)$$

Dans ces conditions, on peut appliquer l'analogue de (I4)

$$(LoT-L)(\phi^{k+p-1}) \geq \alpha \frac{H^2C_v^2}{4C_T^2}\gamma_{k+p} = \alpha \frac{H^2C_v^2}{4C_T^2}\frac{\gamma_{k+p}}{\gamma_k}\gamma_k \geq \alpha \frac{H^2C_v^2}{4C_T^2}(1-\eta)\gamma_k$$

Finalement,

$$L(\phi^{k+p}) - L(\phi^{k+p-1}) \geq \alpha \frac{H^2C_v^2}{4C_T^2}(1-\eta)\gamma_k - O(\gamma_k^{3/2})$$

$$\geq \alpha \frac{H^2C_v^2}{6C_T^2}(1-\eta)\gamma_k \quad \text{si k est assez grand,}$$

et on peut choisir η pour que $1 - \eta > 3/4$, d'où enfin

$$L(\phi^{k+p}) - L(\phi^{k+p-1}) \geq \alpha \frac{H^2C_v^2}{8C_T^2}\gamma_k. \qquad (I12)$$

D'après l'égalité (E) et les inégalités (I12) et (I8) on obtient donc

$$L(\phi^{k+p}) - L(\phi^*) \geq \alpha \frac{H^2C_v^2}{8C_T^2}\gamma_k + 0 - (4C_{DL} + 2C_{DLT})C_v^2\gamma_k$$

$$\geq \{\frac{\alpha H^2}{8C_T^2} - (4C_{DL} + 2C_{DLT})\}C_v^2\gamma_k.$$

Si l'on a pris soin de choisir $H > 1$ assez grand, alors ce minorant est strictement positif et est lui-même minoré par $2 C_{DL} C_v^2 \gamma_k$, quantité à son tour minorée par $2C_{DL} C_v^2 \gamma_{k+p}$, donc par

$2 \sup_{g \in V_{k+p}(\phi^*)} |L(g) - L(\phi^*)|$. Ainsi, on obtient bien $L(\phi^{k+p}) > \sup_{g \in V_{k+p}(\phi^*)} L(g)$. CQFD.

5. Un exemple de comparaison des algorithmes EM, SEM et SAEM

Nous allons comparer le comportement pratique des algorithmes EM, SEM et SAEM pour l'identification d'un mélange gaussien réel à 4 composants dont les caractéristiques sont les suivantes : les 4 proportions sont égales, les 4 moyennes sont $m_1 = 2$, $m_2 = 5$, $m_3 = 9$ et $m_4 = 15$ et les 4 écarts-types sont $\sigma_1 = 0.25$, $\sigma_2 = 0.5$, $\sigma_3 = 1$ et $\sigma_4 = 2$.

Nous avons testé les algorithmes sur deux échantillons, simulés suivant ce mélange, de tailles respectives 100 et 60. Le choix pratique du mode de décroissance vers 0 de la suite (γ_n) de l'algorithme SAEM est bien sûr très important. Nous avons fait de nombreux essais. Dans un souci de concision, nous ne reportons que les résultats de cet algorithme pour deux types de décroissance vers 0 de la suite (γ_n) qui nous semblent significatifs : un mode "lent" ($\gamma_n = \cos(n\Pi/2 \text{ itemax})$), itemax : nombre d'itérations (ici itemax = 200) et un mode "linéaire" ($\gamma_n = -(n + 1) / \text{itemax}$).

Les tableaux suivants résument les résultats obtenus. Dans ces tableaux 'SAEM1' correspond au mode de décroissant lent de γ_n et 'SAEM2' correspond au mode linéaire. Sauf une fois pour l'algorithme EM (colonne EM*), les algoritmes ont été initialisé (avec 5 composants d'une part, et 4 composants d'autre part) à partir d'une position tirée au hasard. Une fois, colonne EM*, l'algorithme EM a été initialisé par les valeurs théoriques des paramètres du mélange simulé. Pour des raisons de simplicité, pour l'algorithme SEM nous ne donnons que les valeurs moyennes de sa loi stationnaire. Enfin, la ligne K init. donne le nombre initial de composants, la ligne K final donne le nombre de composants à la convergence et les lignes 's' donnent les variances des composants.

Echantillon de taille 100

	EM	SEM	SAEM1	SAEM2	EM	SEM	SAEM1	SAEM2
K init.	5	5	5	5	4	4	4	4
K final	5	4	4	5	4	4	4	4
P_1	0.249	0.249	0.249	0.249	0.249	0.249	0.249	0.249
P_2	0.261	0.260	0.258	0.261	0.258	0.258	0.258	0.258
P_3	0.155	0.250	0.255	0.155	0.255	0.255	0.255	0.255
P_4	0.090	0.240	0.238	0.091	0.238	0.238	0.238	0.238
P_5	0.245	---	---	0.245	---	---	---	---
m_1	1.902	1.900	1.902	1.902	1.902	1.902	1.902	1.902
m_2	4.954	4.937	4.938	4.954	9.939	4.939	4.939	4.939
m_3	8.409	9.013	8.980	8.408	8.979	8.979	8.978	8.979
m_4	9.884	14.831	14.912	9.884	14.908	14.906	14.905	14.906
m_5	14.801	---	---	14.800	---	---	---	---
s_1	0.162	0.161	0.162	0.162	0.162	0.162	0.162	0.162
s_2	0.437	0.424	0.414	0.437	0.414	0.414	0.414	0.414
s_3	0.454	0.750	0.800	0.153	0.797	0.796	0.795	0.797
s_4	0.091	4.476	4.011	0.091	4.025	4.031	4.034	4.031
s_5	4.341	---	---	4.343	---	---	---	---

Echantillon de taille 60

	EM	SEM	SAEM1	SAEM2	EM	EM*	SEM	SAEM1	SAEM2
K init.	5	5	5	5	4	4	4	4	4
K final	5	4	4	4	4	4	3	4	4
P_1	0.501	0.250	0.250	0.250	0.502	0.250	0.502	0.250	0.250
P_2	0.045	0.250	0.250	0.248	0.246	0.250	0.240	0.250	0.250
P_3	0.200	0.242	0.243	0.163	0.104	0.242	0.258	0.242	0.242
P_4	0.054	0.258	0.257	0.339	0.148	0.258	---	0.258	0.258
P_5	0.213	---	---	---	---	---	---	---	---
m_1	3.469	2.072	2.072	2.072	3.483	2.072	3.484	2.072	2.072
m_2	8.136	4.853	4.853	4.852	9.741	4.853	9.723	4.853	4.853
m_3	9.892	9.708	9.709	9.866	14.026	9.709	15.027	9.708	9.709
m_4	13.205	15.035	15.038	13.657	15.931	15.034	---	15.033	15.035
m_5	15.323	---	---	---	---	---	---	---	---
s_1	2.044	0.036	0.036	0.036	2.102	0.036	2.105	0.036	0.036
s_2	0.045	0.128	0.128	0.127	0.644	0.128	0.622	0.128	0.128
s_3	0.219	0.650	0.651	0.175	1.987	0.650	4.375	0.650	0.651
s_4	2.836	4.359	4.335	9.749	3.749	4.352	---	4.355	4.347
s_5	4.273	---	---	---	---	---	---	---	---

Commentaires.

Ces simulations mettent en évidence les limites de l'algorithme EM : nécessité de connaître le nombre exact de composants, risque (voir l'échantillon de taille 60) de proposer une solution nettement différente de la solution optimale.

Pour l'algorithme SEM, elles illustrent sa trop grande instabilité pour de petits échantillons, qui peut conduire à sous-estimer le nombre de composants (voir l'échantillon de taille 60, SEM étant initialisé avec 4 composants).

Pour l'algorithme SAEM, le mode lent de décroissance de la suite (γ_n) est incontestablement préférable. Il permet de retrouver le nombre exact des composants, comme le fait l'algorithme SEM pour des échantillons suffisants, mais l'exemple de l'échantillon de taille 60 montre que SAEM conserve cette caractéristique pour de petits échantillons. De plus, il évite mieux que EM la convergence vers une solution erronée. Si l'on ajoute à ces caractéristiques une plus grande simplicité d'exploitation des résultats que pour l'algorithme SEM, il apparaît que l'algorithme SAEM est le mieux adapté des trois à une diffusion dans un logiciel statistique d'usage courant.

ANNEXE

Proposition : il existe a > 0 tel que, pour tout entier j fixé, l'évènement $E_{n,j}$, défini dans 4.2.2. au début du paragraphe "Démonstration du point 1", a lieu pour une infinité de valeurs de n, presque sûrement.

Démonstration :

Nous avons besoin d'un lemme dont la preuve requiert l'usage de l'hypothèse H4.
Notons $w(\phi, e) = < v, V(\phi, e) >$, v étant un vecteur propre de norme 1 associé à une valeur propre $\lambda > 1$ de $DT(\phi^*)$.

Lemme : il existe un réel a > 0 tel que

$$\min+ = \inf_{g \in G} P(w(g, e) > a) > 0$$
$$\min - = \inf_{g \in G} P(w(g, e) > -a) > 0.$$

Preuve du lemme :

Nous allons d'abord voir que l'application qui à ϕ associe la loi de la v.a. $V(\phi, e)$ est continue pour la topologie de la convergence faible. Rappelons d'abord (cf. 4.2.1.), que chaque composant du mélange suit une loi dont la densité est de la forme

$$h(x, a_k) = d(a_k)^{-1} n(x) \exp ({}^t a_k b_k (x)) \text{ pour } k = 1, ..., K.$$

Les coordonnées de $V(\phi, e)$ sont de la forme

$$N^{-1} \sum_{i=1}^{N} (e_k(x_i) - t_k(x_i))$$

pour les proportions p_k et de la forme

$$\{ \sum_{i=1}^{N} e_k(x_i) b_k(x_i) \} / \sum_{i=1}^{N} e_k(x_i) - \{ \sum_{i=1}^{N} t_k(x_i) b_k(x_i) \} / \sum_{i=1}^{N} t_k(x_i)$$

pour les paramètres a_k.

D'après l'hypothèse (H3) les dénominateurs de ces expressions ne s'annulent pas et les $t_k(x_i)$ dépendent continûment de ϕ.
Par suite, pour tout tirage $e = (e(x_i), i=1, ..., N)$, $V(\phi,e)$ dépend continûment de ϕ. Finalement, la continuité de la loi de $V(\phi, e)$ pour la topologie de la convergence faible se déduit de ce que la loi des tirages e dépend continûment de ϕ et de ce que ces tirages sont en nombre fini. Il en résulte que la loi de la v.a. réelle $w(\phi, e)$ dépend continûment de ϕ pour la même topologie.
Nous allons maintenant raisonner par l'absurde. D'après la compacité de G, il existerait pour tout a > 0 (resp. a < 0) une suite (ϕ_m) qui convergerait vers un élément ϕ de G, pour laquelle $P(w(\phi_m, e) > a)$ (resp. < a) convergerait vers 0. Si l'on note F_m (a) la f.d.r. de $w(\phi_m, e)$, on aurait donc convergence de $1 - F_m(a)$ (resp. F_m (a)) vers 0. D'après la continuité faible de la loi de $w(\phi, e)$, il s'ensuivrait que, pour tout point de ce continuité a > 0 (resp. a < 0) de la f.d.r. F

de $w(\phi, e)$, on aurait $1 - F(a) = 0$ (resp. $F(a) = 0$). Il en résulterait que $w(\phi, e) \leq 0$ (resp. ≥ 0) p.s., ce qui contredirait l'hypothèse (H4). CQFD.

Pour démontrer la proposition, nous allons utiliser une version conditionnelle du lemme de Borel-Cantelli, telle qu'elle est donnée dans Hall et Heyde (1980 Corollary 2.3 p.32), à savoir :

soit $\mathcal{F}_n$ une suite croissante de tribus et A_n une suite d'évènement $\mathcal{F}_n$-mesurables ; si $\sum P(A_{n+1} \mid \mathcal{F}_n) = \infty$ p.s. alors A_n a lieu infiniment souvent p.s.

Ici $\mathcal{F}_n$ désignera la tribu engendrée par la suite des tirages $(e^0, ..., e^n)$ et les $E_{n,j}$ prendront la place des A_n ; et en fait, nous allons montrer que, pour tout j, il existe un réel $\delta > 0$ tel que $P(E_{n,j} \mid \mathcal{F}_{n-1}) > \delta$ p.s. pour tout n. Pour cela, nous allons minorer séparément $P(E^+_{n,j} \mid \mathcal{F}_{n-1})$ et $P(E^-_{n,j} \mid \mathcal{F}_{n-1})$.

Rappelons que $q_n = < v, \phi^n - \phi^* >$ et $w_n = < v, V(\phi^n, e^n) >$.

$$P(E^+_{n,j} \mid \mathcal{F}_{n-1}) = \int_{e_n, ..., e_{n+j}} 1_{\{q_n \geq 0\}} 1_{\{w_n \geq a\}} \cdots 1_{\{w_{n+j} \geq a\}} d\, e^n \ldots d\, e^{n+j}$$

$$= 1_{\{q_n \geq 0\}} \int_{e_n, ..., e_{n+j}} 1_{\{w_n \geq a\}} \cdots 1_{\{w_{n+j} \geq a\}} d\, e^n \ldots d\, e^{n+j}$$

car q_n est $\mathcal{F}_{n-1}$ mesurable. Maintenant, par le théorème de Fubini

$$P(E^+_{n,j} \mid \mathcal{F}_{n-1}) \geq 1_{\{q_n \geq 0\}} \int_{e_n} 1_{\{w_n \geq a\}} d\, e^n \left[\ldots \int_{e_{n+j}} 1_{\{w_{n+j} \geq a\}} d\, e^{n+1}\right]$$

Or, d'après le lemme,

$$\int_{e_{n+j}} 1_{\{w_{n+j} \geq a\}} d\, e^{n+j} = P\{e^{n+j} / w(\phi^{n+j}, e^{n+j}) > a\} \geq \min+$$

De même, de proche en proche, on en déduit que

$$P(E^+_{n,j} \mid \mathcal{F}_{n-1}) \geq 1_{\{q_n \geq 0\}} \ (\min+)^{j+1} \text{ p.s.}$$

De manière analogue, on montre que

$$P(E^-_{n,j} \mid \mathcal{F}_{n-1}) \geq 1_{\{q_n \leq 0\}} \ (\min-)^{j+1} \text{ p.s.}$$

Il s'ensuit que

$$P(E_{n,j} \mid \mathcal{F}_{n-1}) \geq [1_{\{q_n \geq 0\}} + 1_{\{q_n \leq 0\}}] \{\inf(\min+, \min-)\}^{j+1} \text{ p.s.}$$
$$P(E_{n,j} \mid \mathcal{F}_{n-1}) \geq \{\inf(\min+, \min-)\}^{j+1} \text{ p.s.}$$

REFERENCES

CELEUX G. et DIEBOLT J. (1986) L'algorithme SEM : un algorithme d'apprentissage probabiliste pour la reconnaissance de mélange de densités. *R.S.A.* **34** n° 2,35-52.

CELEUX G. et DIEBOLT J. (1987) A probabilistie teacher algorithm for iterative maximum likelikood estimation. *Classification and related methods of Data Analysis*. North Holland, 617-623.

CELEUX G. (1987) Thèse d'état. Université Paris 9 Dauphine.

DEMPSTER A.P., LAIRD N.M. et RUBIN D.B. (1977) "Maximum likelikood from incomplete data via the EM algorithm", (with discussion). *J.R.S.S. B.***39**, 1-38.

DIEBOLT J. (1989) Thèse d'état. Université Paris 6.

HALL P. and HEYDE (1988) *Limit theorems for martingales and its applications*. Academic Press, New York.

REDNER R.A. et WALKER H.F. (1984) Mixtures densities, maximum likelikood and the EM algorithm. *SIAM Rev.***26**, 195-249.

TASSI P. (1989) *Méthodes statistiques*. Economica 2ème édition.

TITTERINGTON D.M., SMITH A.F.M., et MAKOV U.E. (1989) *Statistical analysis of finite mixture distribution*. Wiley.

WU C.F.(1983) On the convergence properties of the EM algorithm. *Ann. Statis.* **11**, 95-103.

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique