



HAL
open science

Etude du comportement asymptotique d'un algorithme d'apprentissage probabiliste pour les melanges de lois de probabilite

Gilles Celeux, Jean Diebolt

► **To cite this version:**

Gilles Celeux, Jean Diebolt. Etude du comportement asymptotique d'un algorithme d'apprentissage probabiliste pour les melanges de lois de probabilite. RR-0563, INRIA. 1986. inria-00075991

HAL Id: inria-00075991

<https://inria.hal.science/inria-00075991>

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.

IRIA

CENTRE DE 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° 563

**ÉTUDE DU
COMPORTEMENT ASYMPTOTIQUE
D'UN ALGORITHME
D'APPRENTISSAGE
PROBABILISTE POUR
LES MÉLANGES DE LOIS
DE PROBABILITÉ**

**Gilles CELEUX
Jean DIEBOLT**

Septembre 1986

ETUDE DE COMPORTEMENT ASYMPTOTIQUE D'UN ALGORITHME D'APPRENTISSAGE PROBABILISTE
POUR LES MELANGES DE LOIS DE PROBABILITE

ASYMPTOTICAL BEHAVIOUR OF A PROBABILISTIC TEACHER ALGORITHM
FOR THE MIXTURE PROBLEM

G. CELEUX

INRIA

J. DIEBOLT

CNRS

RESUME

Nous avons développé un algorithme d'apprentissage probabiliste, l'algorithme SEM, pour l'estimation du maximum de vraisemblance des paramètres d'un mélange de lois de probabilité. Jusqu'à présent, l'étude du comportement asymptotique n'avait été menée que dans un cas particulier. Dans cet article, nous étudions ce comportement asymptotique dans le cas général. Cela revient à étudier la loi limite d'un système dynamique à temps discret perturbé par un bruit blanc d'intensité tendant vers zéro. Nos résultats sont de portée assez générale mais nécessitent encore des hypothèses restrictives sur le système dynamique.

MOTS-CLES

Mélange de densité, maximum de vraisemblance, apprentissage probabiliste, chaînes de Markov, système dynamique.

ABSTRACT

Recently we developed a probabilistic teacher algorithm, the SEM algorithm, for the mixture problem. Previously, the asymptotical behaviour had been studied only in a particular case. In this paper, the asymptotical behaviour is studied in the general context. The problem is to find the asymptotical law of a discrete-time dynamical system with small white noise perturbations, the size of which goes to zero. Our results are quite general, but we still need some restrictive assumptions concerning the dynamical system.

KEY-WORDS

Mixture problem, maximum likelihood, probabilistic teacher, Markov chains, dynamical system.

ETUDE DU COMPORTEMENT ASYMPTOTIQUE D'UN ALGORITHME D'APPRENTISSAGE PROBABILISTE
POUR LES MELANGES DE LOIS DE PROBABILITE

1 - INTRODUCTION

Cet article est consacré à l'étude du comportement asymptotique de l'algorithme SEM que nous avons conçu pour le problème de mélanges de distributions de probabilité (cf [CeDi 85], [CeDi 86]).

Ce problème peut s'énoncer ainsi :

Soit un échantillon $E = (x_1, \dots, x_N)$ d'une variable aléatoire X à valeurs dans \mathbb{R}^d , dont la loi est :

$$F = \sum \{p_k F_k \mid k = 1, K\} \text{ où :}$$

Les F_k sont des mesures de probabilité sur \mathbb{R}^d ;

$$\forall k = 1, K \quad 0 < p_k < 1 \text{ et } \sum \{p_k \mid k = 1, K\} = 1$$

p_k est la probabilité qu'un point de l'échantillon suive la loi F_k .

Le problème consiste à estimer le nombre K de composants, les paramètres inconnus p_k ainsi que les lois inconnues F_k .

Ce problème n'a été étudié que dans le cas paramétrique.

Les F_k forment une famille paramétrée que nous noterons

$$F = \{F(\cdot, a), a \in \mathbb{R}^S\}, \text{ l'application } a \rightarrow F(\cdot, a) \text{ étant injective.}$$

De plus, on suppose que les lois $F(\cdot, a)$ admettent une densité $f(\cdot, a)$. Le mélange s'écrit alors :

$$f(x) = \sum \{p_k f(x, a_k) \mid k = 1, K\} \text{ avec :}$$

$$\forall k = 1, K \quad 0 < p_k < 1 \text{ et } \sum \{p_k \mid k = 1, K\} = 1$$

$f(x, a)$ est la densité de probabilité dépendant du paramètre vectoriel a de \mathbb{R}^S .

L'un des algorithmes les plus efficaces pour résoudre ce problème est l'algorithme EM (cf [DLR77], [ReWa84]) dans lequel on cherche à maximiser la vraisemblance de l'échantillon relativement au modèle de mélange par un procédé itératif faisant intervenir deux étapes d'inférence statistique (estimation et maximisation).

L'algorithme SEM, où l'on cherche à supprimer les limitations de l'algorithme EM est fondé sur le même principe, mais adjoint une étape d'apprentissage probabiliste (Stochastique) aux deux étapes précédentes.

Nous présentons ces deux algorithmes de manière détaillée aux paragraphes 2.1 et 2.2.

Tout l'intérêt de l'algorithme SEM réside dans l'étape d'apprentissage probabiliste.

Tout d'abord, elle induit une forme particulière des résultats. Alors que l'algorithme EM fournit une estimation ponctuelle des paramètres du mélange, l'algorithme SEM fournit, à la stationarité, une loi de probabilité sur la suite des estimés de paramètres du mélange. Ce type de convergence en loi présente l'intérêt de fournir des indices pour évaluer le degré d'imbrication des composants du mélange (on analyse la dispersion de la suite des estimés autour de leurs valeurs moyennes).

D'autre part, contrairement à l'algorithme EM, l'utilisation de l'algorithme SEM n'exige pas une connaissance préalable du nombre exact de composants du mélange et fournit des résultats indépendants de la position initiale de l'algorithme.

Il est bien sûr important d'étudier le comportement asymptotique de l'algorithme SEM. Jusqu'à ce jour cette étude n'avait été menée que dans le cas simple d'un mélange à deux composants où seul le paramètre de proportion était inconnu (cf [CeDi 84]).

Nous reprenons ici cette étude dans le cas général. Mais auparavant, nous résumons les résultats obtenus pour le cas simple décrit ci-dessus.

2 - LES RESULTATS DEJA OBTENUS

Le résumé de l'étude sur le comportement asymptotique de l'algorithme SEM lorsqu'un seul paramètre est inconnu va nous permettre de fixer la problématique et les principales notations.

Avant de décrire l'algorithme SEM, nous présentons l'algorithme EM étant donné les forts liens existant entre ces deux algorithmes.

2.1 L'algorithme EM

Cette méthode consiste à résoudre itérativement les équations de vraisemblance ; le Log de la vraisemblance étant :

$$L(x_1, \dots, x_N, a_1, \dots, a_K, p_1, \dots, p_K) = \sum \{ \text{Log} \{ \sum_{k=1, K} p_k f(x_i, a_k) \} \mid i = 0, N \}.$$

A partir d'une solution initiale $(p_k^0, a_k^0 ; k = 1, K)$, l'algorithme est le suivant :

Itération n ($n > 0$)

Etape E (estimation) :

Pour $k = 1, K ; i = 1, N$

$$\text{calcul des } t_k^n(x_i) = p_k^n f(x_i, a_k^n) / \sum \{ p_k^n f(x_i, a_k^n) \mid k = 1, K \}$$

Etape M (maximisation) :

$$\text{Pour } k = 1, K \text{ calcul de } p_k^{n+1} = (1/N) \sum \{ t_k^n(x_i) \mid i = 1, N \}$$

et résolution des équations pour $k = 1, K ; j = 1, s$:

$$\sum (t_k^n(x_i) \partial \text{Log} f(x_i, a_k^{n+1}) / \partial a_{jk} \mid i = 1, N) = 0$$

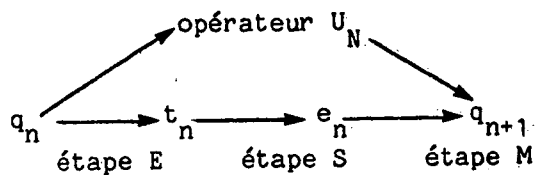
où $a_k = (a_{jk} ; j = 1, s)$.

Notons de même E_N l'opérateur associé à l'étape E et M_N l'opérateur associé à l'étape M.

L'algorithme EM se traduit par l'équation :

$$q_{n+1} = T_N(q_n) = M_N \circ E_N(q_n) = M_N(t_n).$$

L'algorithme SEM peut se schématiser ainsi :



Notons S_N l'opérateur associé à l'étape S et U_N l'opérateur associé à l'algorithme SEM.

L'algorithme SEM se traduit par l'équation :

$$q_{n+1} = U_N(q_n) = M_N \circ S_N \circ E_N(q_n) = M_N \circ S_N(t_n) = M_N(e_n)$$

On peut écrire

$$q_{n+1}(e_0, e_1, \dots, e_n) = U_N(q_n) = M_N(t_n) + [M_N(e_n) - M_N(t_n)]$$

$$q_{n+1}(e_0, e_1, \dots, e_n) = T_N(q_n) + [M_N(e_n) - T_N(q_n)]$$

$$\text{En posant } V_N(q_n, e_n) = M_N(e_n) - T_N(q_n)$$

on obtient

$$\boxed{q_{n+1}(e_0, \dots, e_n) = T_N(q_n) + V_N(q_n, e_n)} \quad (R)$$

La variable aléatoire $V_N(q_n, e_n)$ étant indépendante de $T_N(q_n)$ conditionnellement à q_n .

2.4 Explicitation des deux espaces probabilisés

Il est clair que l'algorithme SEM introduit un deuxième espace probabilisé, celui des tirages aléatoires ; le premier espace probabilisé n'est autre que celui des échantillons.

Il est nécessaire d'utiliser des notations précises pour éviter toute confusion entre ces deux espaces.

2.4.1 L'espace des échantillons

Nous le noterons (Ω, A, P) :

$\Omega = (R^d)^N$, muni de la topologie produit.

A = tribu borélienne associée

$P = \prod_{i \in N} F_i$.

Le i ème élément $X_i(w)$ de l'échantillon désigne la i ème coordonnée de la suite w . Lorsque nous travaillons à échantillon fixé, nous le noterons x_1, \dots, x_N .

Nous utilisons l'indice N systématiquement pour signifier qu'un objet mathématique dépend de l'échantillon. Dans le même esprit, nous parlerons de convergence P - p.s., ou en P - probabilité, etc, lorsque N tend vers l'infini pour signifier que cette convergence est relative à l'espace (Ω, A, P) .

2.4.2 L'espace des tirages aléatoires

A échantillon fixé, l'algorithme SEM utilise, à chaque itération, N tirages indépendants de v.a. multinomiales d'ordre 1 (cf. 2.2).

Nous noterons $(\xi, B(\xi), \pi)$ l'espace probabilisé de ces suites de tirages, dont on notera e l'élément générique.

2.5 Comportement asymptotique de l'algorithme EM dans le cas général

Etant donné la nature des liens existants entre les algorithmes EM et SEM, nous avons besoin d'un résultat sur la convergence de l'algorithme EM quand N tend vers l'infini pour donner un cadre à notre étude.

2.2 L'algorithme SEM

Au départ, on fixe le paramètre K majorant supposé du nombre de composants du mélange et un seuil $c(N)$ compris entre 0 et 1.

Initialisation

En chaque point x_i , $i = 1, N$ on choisit (en général au hasard) les probabilités initiales d'appartenance à l'un des composants :

Soient $t_k^0(x_i)$, $k = 1, K$ avec :

$$0 < t_k^0(x_i) < 1 \text{ et } \sum \{t_k^0(x_i) \mid k = 1, K\} = 1$$

Itération n ($n > 0$) :

Etaps S (Stochastique) :

On tire en chaque point x_i la v.a. multinomiale

$$e^n(x_i) = (e_k^n(x_i) ; k = 1, K)$$

d'ordre un et de paramètres $(t_k^n(x_i) ; k = 1, K)$.

Les réalisations $e^n(x_i)$ définissent une partition $P^n = (P_1^n, \dots, P_K^n)$ de l'échantillon avec :

$$P_k^n = \{x_i / e_k^n(x_i) = 1\}.$$

Si pour un certain k , $\text{card}(P_k^n)$ est plus petit que $Nc(N)$ l'algorithme est re-initialisé.

Sinon :

Etape M (Maximisation) :

On calcule les estimations du maximum de vraisemblance $q_k^{n+1} = (p_k^{n+1}, a_k^{n+1})$ des paramètres du mélange sur la base des sous-échantillons $(P_k^n, k = 1, K)$.

On a :

$$p_k^{n+1} = (1/N) \sum \{e_k^n(x_i) \mid i = 1, N\}$$

L'estimation des a_k^{n+1} dépend bien sûr de la famille paramétrée, posée à priori, des composants du mélange.

Etape E (Estimation) :

A partir des $q_k^{n+1} = (p_k^{n+1}, a_k^{n+1})$, on calcule :

Pour $k = 1, K ; i = 1, N$

$$t_k^{n+1}(x_i) = p_k^{n+1} f(x_i, a_k^{n+1}) / \sum \{p_k^{n+1} f(x_i, a_k^{n+1}) \mid k = 1, K\}.$$

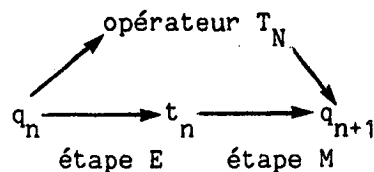
2.3 Rapport formel entre les deux algorithmes

Notons :

$$q_n = (q_k^n, k = 1, K) ; t_n = (t_k^n(x_i), i = 1, N ; k = 1, K) ;$$

$$e_n = (e_k^n(x_i), i = 1, N ; k = 1, K)$$

L'algorithme EM peut se schématiser ainsi :



Notons T_N l'opérateur associé à l'algorithme EM. (L'indice N rappelle la dépendance de cet opérateur relativement à l'échantillon de taille N).

Ce résultat, dû à Redner et Walker [Re Wa 84] se résume ainsi :

Théorème :

Si les densités des composants du mélange appartiennent à la famille exponentielle,

si la matrice d'information de Fisher associée aux paramètres du mélange $I(q)$ est définie positive pour les vraies valeurs des paramètres,

si les vraies proportions p_k sont strictement positives,

alors pour N suffisamment grand, l'unique solution convergente P-p.s. q^N des équations de vraisemblance existe P-p.s. et il existe P-p.s. une norme sur l'espace des paramètres pour laquelle la suite (q_n) des itérés de l'algorithme EM converge vers q^N pour peu que q_0 soit suffisamment proche de q^N , de la manière suivante :

$$\text{il existe } \alpha, \quad 0 \leq \alpha < 1 \text{ tel que } |q_{n+1} - q_n| \leq \alpha |q_n - q^N| \quad \forall n \geq 0.$$

Remarques :

- Ce théorème est d'une grande portée, car la quasi totalité des densités considérées dans les problèmes de mélanges appartiennent à la famille exponentielle.

- Dans ce théorème, on retrouve la condition cruciale pour l'algorithme EM : sa position initiale q_0 doit être proche de la solution optimale q^N qui est inconnue (cf [CeDi 86] pour des illustrations de ce fait).

- Dans un commentaire de leur théorème, Redner et Walker précisent que α sera proche de 0 si les composants du mélange sont bien séparés et que, au contraire, α sera proche de 1 si les composants du mélange sont faiblement séparés. Par là-même, ils cernent bien le problème de la lenteur -pouvant devenir redhibitoire- de l'algorithme EM dans de tels cas.

- L'hypothèse de stricte positivité des proportions p_k rappelle que l'algorithme EM doit disposer du nombre exact de composants.

2.6 Le problème déjà étudié

Dans [CeDi 84] nous avons étudié le comportement asymptotique de l'algorithme SEM dans le cas d'un mélange à deux composants où seul le paramètre p^* ($0 < p^* < 1$) de proportion était inconnu.

Dans ce cas, la formule récursive (R) prend la forme suivante :

$$(R) \quad p_{n+1}^N = T_N(p_n^N) + V_N(p_n^N, e_n) \text{ avec}$$

$$T_N(p_n^N) = \frac{1}{N} \sum_{i=1}^N t^n(x_i) \text{ où}$$

$$t^n(x_i) = p_n^N f(x_i, a_1) / [p_n^N f(x_i, a_1) + (1-p_n^N) f(x_i, a_2)]$$

$$\text{et } V_N(p_n^N, e_n) = \frac{1}{N} \sum_{i=1}^N e^n(x_i) - t^n(x_i) \text{ où}$$

$e_n(x_i)$ est la réalisation d'une v.a. de Bernoulli de paramètre $t^n(x_i)$.

Cette formule récursive (R) peut également s'écrire :

$$p_{n+1}^N = T_N(p_n^N) + \frac{1}{\sqrt{N}} s_N(p_n^N) \eta_{n+1}^N(p_n^N, e_n) \text{ avec :}$$

$$\eta_{n+1}^N(p_n^N, e_n) = \frac{1}{N} \sum_{i=1}^N (e^n(x_i) - t^n(x_i)) \left[\sum_{i=1}^N t^n(x_i) (1-t^n(x_i)) \right]^{-1/2}$$

de telle sorte que :

$E_e(\eta(p, e)) = 0$ et $E_e(\eta^2(p, e)) = 1$ (avec la notation $E_e F = \int_{\mathcal{E}} F(e) \pi(de)$) et avec

$$s_N^2(p_n^N) = \frac{1}{N} \sum_{i=1}^N t^n(x_i) (1-t^n(x_i))$$

de telle sorte que s_N tende vers une limite s différente de 0 lorsque N tend vers l'infini.

Remarquons que $s_N^2(p) = p(1-p) T'_N(p)$ sur $[0,1]$ et donc que $s_N(0) = s_N(1) = 0$.

2.6.2.2 Comportement asymptotique de $p_N, T_N, s_N, \eta^h, r_h$:

Lorsque N tend vers l'infini, p_N tend P-p.s. vers p^* .

Notons $T : [0,1] \rightarrow [0,1]$ la fonction définie par :

$$T(x) = \int_{\mathbb{R}} \frac{x f_1(t) + (1-x) f_2(t)}{x f_1(t) + (1-x) f_2(t)} f^*(t) dt,$$

$f^*(t)$ étant la vraie densité du mélange.

Lorsque N tend vers l'infini, on a P- p.s. :

$$\text{Pour tout } x \in [0,1] \quad \lim T_N(x) = T(x)$$

$$\text{Pour tout } x \in]0,1[\quad \lim T'_N(x) = T'(x) \text{ et } \lim T''_N(x) = T''(x).$$

Il s'en suit que $\lim s_N^2(x) = x(1-x) T'(x)$ et que $\lim r_N = T'(p^*)$ que l'on note r^* .

D'où $\lim s_N^2(p_N) = p^*(1-p^*) r^*$ que l'on note s^* .

Enfin la loi des v.a. sur $\xi \eta^h(x, e)$ converge étroitement, uniformément en $x \in G_h$, vers une loi normale centrée réduite quand h tend vers zéro. En ce sens, la suite $(\eta_n^h ; n \geq 0)$ converge vers un bruit blanc gaussien $(\varepsilon_n(e) ; n \geq 0)$.

2.6.2.3 Représentation de la chaîne $(SEM)_h$

L'étude du comportement asymptotique, quand la taille de l'échantillon tend vers l'infini, de la loi stationnaire ϕ^h de la chaîne $(SEM)_h$ est mal aisée.

Aussi, nous avons été amenés à étudier une autre chaîne, étroitement liée à la chaîne $(SEM)_h$, mais dont l'étude du comportement asymptotique était plus facile, puis à montrer que ces deux chaînes étaient asymptotiquement équivalentes.

Cette chaîne auxiliaire $(V)_h$ est définie ainsi :

$$V_{n+1}^h = t_h(V_n^h) + h^{-\frac{1}{2}} s_n(V_n^h) \eta_{n+1}^h(V_{n,e}^h),$$

les fonctions T_h et s_h étant modifiées de manière à coïncider sur G_h avec les fonctions T_h et s_h définies précédemment et à être constantes à l'extérieur de G_h ; on note ainsi ces constantes :

$$T_h(x) = t_h \in G_h \quad \text{si } x \notin G_h$$

et

$$s_h(x) = v_h \geq 0 \quad \text{si } x \notin G_h.$$

Ainsi la différence entre les chaînes $(SEM)_h$ et $(V)_h$ réside dans leurs comportements respectifs lorsque X sort de G_h .

- Pour $(SEM)_h$: si $X_{n+\frac{1}{2}}^h \notin G_h$, la v.a. X_{n+1}^h est tirée suivant une loi γ_h fixée à l'avance.

- Pour $(V)_h$: si $V_{n+1}^h \notin G_h$, la v.a. V_{n+2}^h est tirée suivant une loi normale de moyenne t_h et d'écart-type v_h strictement positif, ou selon la mesure de Dirac au point t_h si $v_h = 0$.

On observe donc un décalage d'indice à chaque excursion de X hors de G_h .

La chaîne $(V)_h$ est ergodique et admet donc une loi stationnaire μ^h .

Par ailleurs, on considère dans la suite les chaînes normalisées $(Y_n^h ; n \geq 0)$ et $(W_n^h ; n \geq 0)$ définies par :

$$Y_n^h = h^{-\frac{1}{2}} X_n^h \quad \text{et} \quad W_n^h = h^{-\frac{1}{2}} V_n^h$$

et de lois stationnaires respectives ψ^h et ν^h , ceci en vue d'énoncer des théorèmes limites.

2.6.2.4 Les théorèmes de convergence

Le premier théorème décrit le comportement asymptotique de la loi stationnaire ν^h de la chaîne $(V)_h$ normalisée.

2.6.1 Notations

On pose $h = \frac{1}{N}$. Dans toute la suite, pour alléger les calculs, on utilisera des variables et des fonctions centrées.

Ainsi dans le cas présent, on considère :

$$X_n^h = P_n^N - P_N$$

et pour $x \in [-p_N, 1-p_N] = G$

$$T_h(x) = T_N(x+p_N) - p_N$$

$$s_h(x) = s_N(x+p_N)$$

$$\eta_{n+1}^h(x, e) = \eta_{n+1}^N(x + p_N, e)$$

$$r_h = T_h'(0) = T_N'(P_N)$$

$$G_h = [c(N) - P_N, (1-c(N)) - p_N]$$

où

- P_N est l'unique point fixe sur $]0,1[$ de T_N . Ce point fixe est attractif et réalise le maximum global de la vraisemblance sur $[0,1]$ (cf [Ced 84]).

- P_n^N est l'estimation du paramètre p^* à la $n^{\text{ième}}$ itération de l'algorithme SEM.

L'algorithme SEM est re-initialisé (cf 2.2) dans le cas où $T_N(P_n^N) + V_N(P_n^N, e^N)$ n'appartient pas à l'intervalle $[c(N), 1-c(N)]$:

On procède alors au tirage au hasard de p_{n+1}^N selon une loi de probabilité Y_N à support contenu dans $[c(N), 1-c(N)]$ et fixée à l'avance (à échantillon donné).

Dans ce cas, il est commode de noter

$$X_{n+\frac{1}{2}}^h = T_h + V_h$$

D'où l'écriture de l'algorithme :

$$\text{Si } X_{n+\frac{1}{2}}^h \in G_n, X_{n+1}^h = X_{n+\frac{1}{2}}^h$$

Si $X_{n+\frac{1}{2}}^h \notin G_n$, $X_{n+1}^h = \Gamma_{n+1}$: une réalisation (indépendante des précédentes) de la loi de relance γ .

Remarquons que l'intervalle G_n ne contient zéro que si $p_N \in [c(N), 1-c(N)]$ pour N assez grand. Ceci est vérifié si $c(N)$ tend vers zéro quand N tend vers l'infini, mais ce n'est pas forcément le cas si $c(N) = \text{Cste}$.

On introduit, de plus, la quantité $R(h) = \inf_{x \in G_h} \frac{|T_h(x)|}{|x|}$

Cette quantité est bien définie car $T_h(0) = 0$ et T_h est dérivable en 0 ; on a :

$$0 < R(h) < 1 \quad \text{et} \quad \lim_{h \rightarrow 0} R(h) = 1 \quad \text{si} \quad \lim_{h \rightarrow 0} c(h) = 0$$

(cf [CeDi 84]).

2.6.2 Résumé des résultats [CeDi 84]

Dans cette partie, nous reprenons la démarche qui nous a conduit aux théorèmes de convergence présentés au paragraphe 2.6.2.4. En effet, l'étude du cas général sera de même nature. Tous les résultats énoncés sont démontrés dans [CeDi 84].

2.6.2.1 Ergodicité (à h fixé) de la chaîne $(X_n^h ; n \geq 0)$

A h fixé, la suite $(X_n^h ; n \geq 0)$ des itérés de l'algorithme SEM constitue une chaîne de Markov ergodique sur l'espace $(\xi, B(\xi), \pi)$. Elle admet donc une loi stationnaire que nous notons ϕ^h .

On désignera par $(SEM)_h$ cette chaîne indexée par h .

Le deuxième théorème montre qu'asymptotiquement l'écart entre les lois stationnaires ψ^h et ν^h est faible.

Théorème 1 :

Si $c(h)$ tend vers 0 assez lentement quant h tend vers zéro, la loi limite de ν^h est une loi normale $N(0, \sigma)$ avec :

$$\sigma^2 = \frac{s^*}{(1-p^{*2})} \text{ strictement positif.}$$

Théorème 2

La variation totale de la mesure signée $\psi^h - \nu^h$ tend vers 0 quant h tend vers zéro.

Il ressort de ces deux théorèmes que, à l'état stationnaire, lorsque N tend vers l'infini, la suite des estimés p_n^N par l'algorithme SEM de la proportion p^* peut être considéré comme une suite de réalisation indépendante d'une loi normale de moyenne p_N et de variance $\frac{\sigma^2}{N}$ ($= h \sigma^2$).

III - ETUDE DU CAS GENERAL

Le but de cet article est de présenter dans le cas général où tous les paramètres du mélange sont inconnus des résultats analogues à ceux décrits dans le paragraphe 2.6.

Pour ce faire, il nous faut reprendre les notations et les définitions et préciser les hypothèses des théorèmes de convergence asymptotique que nous démontrons aux paragraphes 3.2 et 3.3.

3.1 Définitions, notations et hypothèses

Soit $q = (p_1, \dots, p_{K-1}, a_1, \dots, a_K) \in \mathbb{R}^p$ avec $p = K-1 + sK$ le paramètre vectoriel à estimer.

Soit q^N la solution asymptotiquement convergente de l'algorithme EM sous les hypothèses du théorème énoncé au paragraphe 2.5, à savoir :

(H1) : les densités du mélange appartiennent à la famille exponentielle.

(H2) : La matrice d'information de Fisher $I(q)$ associée aux paramètres du mélange est définie positive pour la vraie valeur de q .

(H3) : $\forall k = 1, K \ p_k > 0$.

Rappelons que l'on a posé $h = \frac{1}{N}$, N étant la taille de l'échantillon.

Soit T_h , fonction de \mathbb{R}^p dans \mathbb{R}^p , l'opérateur de l'algorithme EM.

Pour $h > 0$, on considère une famille décroissante $(G_h ; h > 0)$ de boréliens de \mathbb{R}^p tels que :

(H4) Soit $G_h = G$, soit $\bigcup_{h>0} G_h = G$, borélien fixe de \mathbb{R}^p .

(H5) Il existe un réel b , $0 < b \leq 1$, tel que la boule B de \mathbb{R}^p de centre 0 et de rayon b soit contenue dans tous les boréliens G_h pour h assez petit.

On considère des réels $R(h)$, $0 \leq R(h) < 1$ tels que :

(H6) Soit $R(h) = R$ pour tout $h > 0$, avec $0 < R < 1$; soit $\lim_{h \rightarrow 0} R(h) = 1$, la fonction $R(h)$ étant décroissante.

On suppose que les fonctions $T_h(x)$ sont telles que :

(H7) $\forall x \in G_h, |T_h(x)| \leq R(h) |x|$

où $|\cdot|$ désigne la norme sur \mathbb{R}^p introduite par Redner et Walker (voir le théorème du paragraphe 2.5)

(H8) Pour $x \notin G_h$, on a $T_h(x) = t_h$ où t_h est une constante appartenant à $R(h) G_h$ homothétique de G_h dans le rapport $R(h)$.

(H9) Il existe un réel q , $0 < q < 1$, une matrice $r \in M_p(\mathbb{R})$ (ensemble des matrices carrées de dimension p à valeurs réelles) et, pour tout $h > 0$, une matrice $r_h \in M_p(\mathbb{R})$ tels que :

(3.1.1) $\forall h > 0 \quad \|r_h\| \leq q$ où $\|\cdot\|$ désigne la norme d'opérateur associée à la norme introduite dans (H7)

(3.1.2) $\lim_{h \rightarrow 0} r_h = r$

(3.1.3) $x \in B \implies |T_h(x) - r_h(x)| \leq Cste |x|^2$

Remarques

- (3.1.3) généralise le cas où T_h est de classe C^2 avec une différentielle seconde bornée sur \mathbb{R}^p .

- De (3.1.3) on tire que pour tout réel α compris entre 0 et 1, on a $|T_h(x) - r_h(x)| < |x|^{1+\alpha}$ pour $x \in B$, car alors $|x| < 1$.

Nous allons maintenant préciser, dans le cas général, la forme de la chaîne de Markov homogène ergodique $(SEM)_n$ associé à la suite des itérés $(X_n^h; n \geq 0)$ de l'algorithme SEM.

Remarque :

L'ergodicité de cette chaîne peut se prouver de manière analogue au cas traité dans [CeDi 84] page 30.

Considérons des applications à valeurs matricielles

$$s_h : \mathbb{R}^p \rightarrow m_p(\mathbb{R}) \text{ telles que :}$$

(H10) Pour tout $h > 0$, pour tout $x \in G_h$, $s_h(x) = v_h$ matrice fixe.

(H11) On a (3.1.4) $\sup (\|s_h(x)\|, h > 0, x \in \mathbb{R}^p) = C$

avec $C < \infty$.

(H12) Il existe une matrice $s \in m_p(\mathbb{R})$ telle que :

$$(3.1.5) \lim_{h \rightarrow 0} s_h(0) = s$$

$$(3.1.6) x \in B \implies \|s_h(x) - s_h(0)\| \leq Cste |x|$$

Remarque :

De la même manière que pour (3.1.3), il s'en suit que pour $x \in B$, pour tout α compris entre 0 et 1, on a :

$$\|s_h(x) - s_h(0)\| \leq Cste |x|^\alpha$$

On introduit, pour n entier positif, $h > 0$, $x \in \mathbb{R}^p$ des v.a. sur $(\xi, B(\xi), \pi)$, à valeurs dans \mathbb{R}^p , $\eta_n^h(x, e)$ et $\varepsilon_n(e)$, les $\varepsilon_n(e)$ étant des v.a. gaussiennes, telles que :

(H13) A $h > 0$ et $x \in \mathbb{R}^p$ fixés, les $\eta_n^h(x, \cdot)$ (resp. les $\varepsilon_n(\cdot)$) sont indépendantes et identiquement distribuées.

(H14) On a :

$$(3.1.7) E_e(\eta_n^h(x,e)) = E_e(\epsilon_n(e)) = 0$$

$$(3.1.8) E_e((\eta_n^h(x,e))^2) = E_e(\epsilon_n^2(e)) = 1$$

(3.1.9) loi $_e(\eta_n^h(x,e))$ converge étroitement, uniformément en x ,

vers loi $_e(\epsilon_n(e))$ lorsque h tend vers 0.

Remarque

(3.1.9) s'appuie sur une généralisation directe de la proposition 4 de [CeDi 84].

La chaîne de Markov (SEM) $_h$ peut s'écrire ainsi :

$$\text{Soit } X_{n+\frac{1}{2}}^h = T_h(X_n^h) + h^{\frac{1}{2}} s_h(X_n^h) \eta_{n+1}^h(X_n^h; e)$$

$$(3.1.10) X_{n+1}^h = X_{n+\frac{1}{2}}^h \text{ si } X_{n+\frac{1}{2}}^h \in G_h$$

$X_{n+1}^h = \Gamma_{n+1}$ sinon, où Γ_{n+1} est une réalisation tirée indépendamment des v.a. X_0^h, \dots, X_n^h et de $\eta_{n+1}^h(X_n^h, \cdot)$, selon une loi de probabilité γ_h , à support contenu dans le borélien $R(h) G_h$.

La chaîne de Markov homogène ergodique (V) $_h$ peut alors s'écrire :

$$(3.1.11) V_{n+1}^h = T_h(V_n^h) + h^{\frac{1}{2}} s_h(V_n^h) \eta_{n+1}^h(V_n^h, e)$$

où, rappelons-le :

$$T_h(x) = t_h \text{ si } x \notin G_h \text{ et } s_h(x) = v_h \text{ si } x \in G_h.$$

Rappelons aussi (cf paragraphe 2.6) que la différence entre les deux chaînes (SEM) $_h$ et (V) $_h$ réside dans le décalage d'indice se produisant à chaque excursion à l'extérieur de G_h .

On suppose :

(H15) X_0^h et $\eta_1^h(x, \cdot)$ sont indépendantes.

V_0^h et $\eta_1^h(x, \cdot)$ sont indépendantes.

(H16) $E_e((X_0^h)^2) < \infty$; $E_e((V_0^h)^2) < \infty$

(H17) support (loi X_0^h) et support (loi (V_0^h)) $\subset G_h$.

On notera la loi stationnaire de la chaîne $(SEM)_h$ (resp. $(V)_h$) ϕ^h (resp. μ^h). De plus, on note :

$$(3.1.12) \phi_n^h = \text{loi}(X_n^h), \mu_n^u = \text{loi}(V_n^h)$$

$$(3.1.13) \psi_n^h = \text{loi}(Y_n^h) \text{ où } Y_n^h = h^{-1/2} X_n^h$$

$$v_n^h = \text{loi}(W_n^h) \text{ où } W_n^h = h^{-1/2} V_n^h$$

(3.1.14) ψ^h et v^h dénotent les lois stationnaires des chaînes $(Y_n^h ; n \geq 0)$ et $(W_n^h ; n \geq 0)$ respectivement.

On utilisera dans la démonstration du théorème du paragraphe 3.2 deux chaînes de Markov auxiliaires : la première $(Z_n^h ; n \geq 0)$ définie par :

$$(3.1.15) Z_{n+1}^h = r_h Z_n^h + s_h(0) \epsilon_{n+1}(e) \text{ et } Z_0^h = Y_0^h$$

Il s'agit d'une chaîne auto-régressive linéaire d'ordre un avec un bruit blanc gaussien et on supposera que pour h assez petit :

(H18) La chaîne $(Z_n^h ; n \geq 0)$ est ergodique.

La deuxième $(Z_n ; n \geq 0)$ est définie par :

$$(3.1.16) Z_{n+1} = r Z_n + s \epsilon_{n+1}(e)$$

Cette chaîne est de même nature que la précédente et l'on suppose de même que :

(H19) La chaîne $(Z_n ; n \geq 0)$ est ergodique et on notera λ sa mesure de probabilité stationnaire.

λ est donc la loi gaussienne sur R^p centrée et de matrice variance régulière (d'après (H19)) définie comme la somme de la série (convergente d'après (H9))

$\sum_{l=0}^{\infty} r^l s s' (r')^l$ où s' et r' désignent les matrices transposées de r et de s .

Remarque :

Les hypothèses (H18) et (H19) d'ergodicité des chaînes auxiliaires ne sauraient être prouvées directement dans le cas général car l'explication de r_h , $s_h(0)$, r et s conduiraient à des calculs inextricables. Elles nous apparaissent réalistes.

3.2 Loi limite de la loi stationnaire de $(W)_h$

Théorème 1 :

Sous les hypothèses (H1) à (H19) et si

(3.2.1) $\lim_{h \rightarrow 0} h^a (1-R(h))^{-1} = 0$ avec $a = \frac{\alpha}{2(1+\alpha)}$ et α quelconque dans $]0,1[$
alors :

(3.2.2) $\lim_{h \rightarrow 0} \text{étroite } v^h = \lambda$

où λ est la mesure de probabilité stationnaire de la chaîne $(Z_n ; n \geq 0)$ définie en (3.1.16).

Plan de la démonstration

En raison de l'hypothèse (H14), nous allons pouvoir utiliser le théorème de représentation de Skorokhod (cf [Bre 68], section 13.9), selon lequel il est possible de définir, sur un espace probabilisé adéquat $(W, B(W), m)$, des v.a.

$\hat{\eta}_n^h(x, w)$ et $\hat{\varepsilon}_n(w)$ de mêmes lois respectivement que $\eta_n^h(x, e)$ et $\varepsilon_n(e)$ et telles que la limite presque-sûre pour la mesure m quand h tend vers 0 de $\hat{\eta}_n^h(x, \cdot)$ est $\hat{\varepsilon}_n(\cdot)$ P-p.s., uniformément en $x \in G_h$ et en $n \geq 0$.

Le principe de la démonstration sera alors de prouver des résultats de convergence dans l'espace $L^{1+\beta}(W)$ (et non $L^{1+\beta}(\xi)$) avec $0 < \beta < 1$. Ces résultats pourront ensuite être retraduits pour les v.a. initiales définies sur ξ et fourniront alors des résultats de convergence en loi.

Pour simplifier les notations, nous omettrons toute référence explicite aux v.a. dans W : les raisonnements des étapes 2 à 5 porteront sur les v.a. de W sans que nous le signalions ; ainsi nous noterons abusivement $\eta_n^h(x, e)$ au lieu de $\hat{\eta}_n^h(x, w)$ et $\varepsilon_n(e)$ au lieu de $\hat{\varepsilon}_n(w)$.

La dernière étape (étape 6) sera consacrée à la déduction des résultats de convergence en loi des v.a. définies sur ξ .

Ceci étant posé, le plan de la démonstration est le suivant :

L'étape 1 est consacrée à l'établissement d'inégalités utiles pour la suite (Lemmes 1 et 2).

L'étape 2 est consacré à la majoration dans $L^{1+\beta}$ ($0 < \beta < 1$) de la norme de la différence entre la v.a. W_n^h est la première v.a. auxiliaire Z_n^h (Lemmes 3 et 4).

L'étape 3 est consacrée à la majoration dans $L^{1+\beta}$ de la norme de la différence entre la v.a. W_n^h et la deuxième v.a. auxiliaire Z_n (Lemme 5).

L'étape 4 est consacrée à la majoration dans $L^{1+\beta}$ d'une v.a. V^h dont la loi est la loi stationnaire de la chaîne $(V_n^h ; n > 0)$ (Lemme 6).

Dans l'étape 5, on montre que W_n^h converge vers Z_n dans $L^{1+\beta}$ lorsque h tend vers zéro.

L'étape 6 est consacrée à la traduction des résultats précédents pour obtenir le résultat de convergence en loi annoncé.

Etape 1Lemme 1 :

Pour tout γ ($1 < \gamma < 2$), pour tout x , tout $h > 0$, tout $n \geq 0$, on a l'inégalité :

$$(3.2.3) \quad E^{1/\gamma} |W_n^h|^\gamma \leq C(1-R(h))^{-1} + h^{-\frac{\gamma}{2}} R^n(h) v(\gamma; h)$$

où $C = \sup_{h>0; x} ||s_n(x)||$ et où on a défini

$$(3.2.4) \quad v(\gamma, h) = E^{1/\gamma} |V_0^h|^\gamma$$

Remarquons que $v(\gamma, h) \leq E^{\frac{\gamma}{2}} |V_0^h|^2 < \infty$ d'après l'annexe 1 et (H16).

Démonstration : Prouvons (3.2.3)

$$V_{n+1}^h = T_h(V_n^h) + h^{\frac{\gamma}{2}} s_h(V_n^h) \eta_{n+1}^h(V_n^h, e) \text{ par (3.1.11)}$$

$$|V_{n+1}^h| \leq R(h) |V_n^h| + h^{\frac{\gamma}{2}} C \eta_{n+1}^h(V_n^h, e) \text{ d'après (H7) et (H11).}$$

D'où par l'inégalité de Minkowski, $\forall \gamma$ ($1 < \gamma < 2$) :

$$(3.2.5) \quad E^{1/\gamma} |V_{n+1}^h|^\gamma \leq R(h) E^{1/\gamma} |V_n^h|^\gamma + h^{\frac{\gamma}{2}} C E^{1/\gamma} (|\eta_{n+1}^h(V_n^h, e)|^\gamma)$$

et d'autre part, on a :

$$E^{1/\gamma} (|\eta_{n+1}^h(V_n^h, e)|^\gamma) \leq E^{\frac{\gamma}{2}} (|\eta_{n+1}^h(V_n^h, e)|^2) \text{ d'après l'annexe 1.}$$

Or V_n^h est une v.a. indépendante du $(n+1)^{\text{ième}}$ tirage (que nous notons e). La loi du couple (V_n^h, e) est donc le produit de la loi μ_n^h de V_n^h et de la loi multinomiale du $(n+1)^{\text{ième}}$ tirage. On peut appliquer le théorème de Fubini et la relation (3.1.8) :

$$E(|\eta_{n+1}^h(V_n^h, e)|^2) = \int E_e |\eta_{n+1}^h(x, e)|^2 \mu_n^h dx = \int 1 \mu_n^h dx = 1$$

Posons $u_n = E^{1/Y} |V_n^h|^Y$ et donc $u_0 = v(\gamma, h)$, de l'inégalité (3.2.5) et du calcul précédent on tire :

$$u_{n+1} \leq R(h) u_n + h^{\frac{Y}{2}} C / (1-R(h))$$

Or, on a défini $W_n^h = h^{-\frac{Y}{2}} V_n^h$ d'où l'inégalité (3.2.3). \square

Lemme 2 :

Pour tout $x = h^{\frac{Y}{2}} y$ et tout α ($0 < \alpha \leq 1$), on a les inégalités suivantes :

$$(3.2.6) \quad |h^{-\frac{Y}{2}} T_h(h^{\frac{Y}{2}} y) - r_h y| \leq \text{Cste } h^{\alpha/2} |y|^{1+\alpha}$$

$$(3.2.7) \quad ||s_h(h^{\frac{Y}{2}} y) - s_h(0)|| \leq \text{Cste } h^{\alpha/2} |y|^\alpha$$

Démonstration :

Si $x = h^{\frac{Y}{2}} y \in B$ ces inégalités se déduisent immédiatement de (3.1.3) et (3.1.6).

Si $x = h^{\frac{Y}{2}} y \notin B$, on a $|y| > b h^{-\frac{Y}{2}}$ d'où :

$$|h^{-\frac{Y}{2}} T_h(h^{\frac{Y}{2}} y)| < R(h) |y| \leq |y|^{1+\alpha} |y|^{-\alpha} \leq b^{-\alpha} h^{\alpha/2} |y|^{1+\alpha}$$

et de manière analogue

$$||r_h y|| < b^{-\alpha/2} h^{\alpha/2} |y|^{1+\alpha} \text{ ce qui prouve que (3.2.6) est vraie}$$

partout.

$$\text{Et } ||s_h(0)|| \leq \text{Cste } \frac{|y|^\alpha}{|y|^\alpha} \leq \text{Cste } b^{-\alpha} h^{\alpha/2} |y|^\alpha$$

Il en est de même pour $||s_h(h^{\frac{Y}{2}} y)||$ et donc (3.2.7) est vraie partout. \square

Etape 2 : Comparaison de W_n^h et Z_n^h

Rappelons qu'à partir de maintenant et jusqu'à l'étape 5 incluse, nous allons travailler à partir de v.a. $\hat{\eta}_n^h(x, w)$ et $\hat{\varepsilon}_n(w)$ définies sur un espace

probabilisé $(W, B(W), m)$ et construites à partir de $\eta_n^h(x, e)$ et $\epsilon_n(e)$ par le théorème de représentation de Skorohod. $\hat{\eta}_n^h$ et $\hat{\epsilon}_n$ induisent les chaînes $\hat{V}_n^h, \hat{W}_n^h, \hat{Z}_n^h, \hat{Z}_n$ au moyen des formules définies au paragraphe 3 pour les chaînes V_n^h, W_n^h, Z_n^h et Z_n à l'aide de η_n^h et ϵ_n avec :

$$\text{loi}(\hat{V}_0^h) = \text{loi}(V_0^h) \text{ et } \text{loi}(\hat{Z}_0^h) = \text{loi}(Z_0^h).$$

Nous insistons sur le fait que, dans la suite, pour alléger les notations, nous confondrons dans l'écriture les v.a. définies sur W et les v.a. définies sur ξ .

Nous allons montrer :

Lemme 3 :

Soit $u_n^h = E^{1/1+\beta} |W_n^h - Z_n^h|^{1+\beta}$ et donc $u_0^h = 0$, où β est un réel strictement compris entre 0 et 1.

Alors on a l'inégalité :

$$(3.2.15) \quad u_n^h \leq \text{Cste} [C h^{\frac{\alpha}{2(1+\alpha)}} (1-R(h))^{-1} + h^{-\frac{1}{2(1+\alpha)}} v(\gamma; h)]^{1+\alpha} \\ + \text{Cste} [C h^{\frac{\gamma}{2}} (1-R(h))^{-1} + v(\delta; h)]^\alpha + a(h)$$

avec $\lim_{h \rightarrow 0} a(h) = 0$ et où α est un réel de $]0, 1]$, δ est un réel lié à α , compris strictement entre 1 et 2.

Démonstration

Des définitions de W_n^h et Z_n^h , on tire :

$$W_{n+1}^h - Z_{n+1}^h = r_h (W_n^h - Z_n^h) + h^{-\frac{\gamma}{2}} T_h(h^{\frac{\gamma}{2}} W_n^h) - r_h W_n^h \\ + s_h(0) [\eta_{n+1}^h(h^{\frac{\gamma}{2}} W_n^h, e) - \epsilon_{n+1}(e)] \\ + (s_h(h^{\frac{\gamma}{2}} W_n^h) - s_h(0)) \eta_{n+1}^h(h^{\frac{\gamma}{2}} W_n^h, e)$$

En appliquant les inégalités (3.2.6) et (3.2.7) il vient pour tout α de]0,1[

$$(3.2.8) \quad |W_{n+1}^h - Z_{n+1}^h| \leq r_n |W_n^h - Z_n^h| + \text{Cste} |W_n^h|^{1+\alpha} \\ + \text{Cste} h^{\alpha/2} |W_n^h|^\alpha |\eta_{n+1}^h| (h^{\frac{1}{2}} W_n^h, e) \\ + \text{Cste} |\eta_{n+1}^h| (h^{\frac{1}{2}} W_n^h, e) - \varepsilon_{n+1}(e)$$

Soit β un nombre réel arbitraire strictement compris entre 0 et 1, on peut choisir trois nombre réels α , p et q vérifiant les conditions suivantes (cf annexe 2) :

$$(3.2.9) \quad 0 < \alpha < 1 ; (1+\alpha)(1+\beta) < 2 ; p, q > 1 ; \frac{1}{p} + \frac{1}{q} = 1 ; q(1+\beta) < 2 ; \\ 1 < p\alpha(1+\beta) < 2$$

Notons $u_n^h = E^{1/(1+\beta)} |W_n^h - Z_n^h|^{1+\beta}$ et donc $u_0^h = 0$.

En appliquant l'inégalité de Minkowski à (3.2.8) on obtient :

$$(3.2.10) \quad u_{n+1}^h \leq r_n u_n^h + \text{Cste} A_n + \text{Cste} B_n + \text{Cste} C_n$$

avec :

$$A_n = h^{\alpha/2} E^{1/1+\beta} |W_n^h|^{(1+\alpha)(1+\beta)} \\ B_n = h^{\alpha/2} E^{1/1+\beta} (|W_n^h|^\alpha)^{(1+\beta)} |\zeta_{n+1}|^{1+\beta}$$

où on a posé

$$\zeta_n = \eta_n^h (h^{\frac{1}{2}} W_n^h, e) \\ C_n = E^{1/1+\beta} |\zeta_{n+1} - \varepsilon_{n+1}(e)|^{1+\beta}$$

En remarquant que $u_0^h = 0$ et que $||r^h|| \leq q \leq 1$ d'après (3.1.1), on tire aisément de (3.2.10) que :

$$(3.2.11) \quad u_n^h \leq \text{Cste} \sum_{\ell=0}^{n-1} q^{n-1-\ell} A_\ell + \text{Cste} \sum_{\ell=0}^{n-1} q^{n-1-\ell} B_\ell + \text{Cste} \sum_{\ell=0}^{n-1} q^{n-1-\ell} C_\ell$$

Par ailleurs, on a

$$A_\ell = h^{\alpha/2} (E^{1/(1+\alpha)(1+\beta)} |W_\ell^h|^{(1+\alpha)(1+\beta)})^{1+\alpha} \\ \leq h^{\alpha/2} [c(1-R(h))^{-1} h^{-\frac{\gamma}{2}} R^\ell(h) v(\gamma, h)]^{1+\alpha} \text{ d'après (3.2.3)}$$

avec $\gamma = (1+\alpha)(1+\beta)$

D'où

$$(3.2.12) \quad A_\ell \leq [C h^{\alpha/2(1+\alpha)} (1-R(h))^{-1} + h^{1/2(1+\alpha)} v(\gamma; h)]^{1+\alpha}$$

puisque $R^\ell(h) < 1$.

D'autre part, d'après le choix fait pour α , p et q (cf (3.2.9)) on peut appliquer l'inégalité de Hölder à B_ℓ , il vient :

$$B_\ell \leq h^{\alpha/2} E^{1/p(1+\beta)} ||W_n^h||^{p\alpha(1+\beta)} E^{1/q(1+\beta)} ||\zeta_{n+1}||^{q(1+\beta)}$$

Puisque $q(1+\beta)$ et $p\alpha(1+\beta)$ sont inférieurs à 2, on en déduit (cf Annexe 1 et (3.2.3))

$$B_\ell \leq [C h^{\frac{\gamma}{2}} (1-R(h))^{-1} + R^\ell(h) v(p\alpha(1+\beta); h)]^\alpha \text{ d'où} \\ (3.2.13) \quad B_\ell \leq [C h^{\frac{\gamma}{2}} (1-R(h))^{-1} + v(\delta; h)]^\alpha \text{ avec } \delta = p\alpha(1+\beta) \text{ et } 1 < \delta < 2$$

Enfin nous allons voir que

$$(3.2.14) \quad C_n = E^{1/(1+\beta)} ||\zeta_{n+1}^{-\varepsilon_{n+1}}(e)||^{1+\beta}$$

tend vers 0, uniformément en n , lorsque $h \rightarrow 0$.

En appliquant alors (3.2.12), (3.2.13) et (3.2.14) à (3.2.11), il viendra aisément :

$$\begin{aligned}
u_n^h &\leq \frac{\text{Cste}}{1-q} [C h^{\alpha/2(1+\alpha)} (1-R(h))^{-1} + h^{-1/2(1+\alpha)} v(\gamma;h)]^{1+\alpha} \\
&+ \frac{\text{Cste}}{1-q} [C h^{\frac{\gamma}{2}}(1-R(h))^{-1} + v(\delta;h)]^\alpha \\
&+ a(h) \quad \text{avec } \lim_{h \rightarrow 0} a(h) = 0
\end{aligned}$$

Soit l'inégalité :

$$\begin{aligned}
(3.2.15) \quad u_n^h &\leq \text{Cste} [C h^{\alpha/2(1+\alpha)} (1-R(h))^{-1} + h^{-\alpha/2(1+\alpha)} v(\gamma;h)]^{1+\alpha} \\
&+ \text{Cste} [C h^{\frac{\gamma}{2}}(1-R(h))^{-1} + v(\delta;h)]^\alpha + a(h)
\end{aligned}$$

avec $\lim_{h \rightarrow 0} a(h) = 0$.

Pour prouver (3.2.14), nous allons démontrer ce résultat pour $E^{\frac{\gamma}{2}} \left| \left| \zeta_{n+1} - \varepsilon_{n+1}(e) \right| \right|^2$ car il est plus commode de raisonner dans l'espace de Hilbert L^2 , et (3.2.14) se déduira de l'annexe 1 puisque $1 < 1+\beta < 2$.

Ce résultat est l'objet du lemme 4 appliqué aux v.a. ζ_{n+1} , ε_{n+1} et V_n^h :

Lemme 4 :

Soient les v.a. à valeurs dans R^d $U^h(x,e)$ et $V(e)$ du second ordre, avec $x \in R^d$, $h > 0$; soit $X^h(e)$ une v.a. à valeurs dans R^d , de loi ϕ^h pour chaque $h > 0$.

Supposons que :

$$(i) \quad \lim_{h \rightarrow 0} E | U^h(x,.) |^2 = E | V |^2 \text{ uniformément en } x$$

$$(ii) \quad \lim_{h \rightarrow 0} \text{p.s. } U^h(x,.) = V \text{ uniformément en } x$$

$$\text{Alors } E | U^h(X^h(e);e) - V(e) |^2 \rightarrow 0 \text{ quand } h \rightarrow 0.$$

Démonstration :

Par (i), la famille des v.a. $U^h(x,.)$ est équi-intégrable, uniformément en

x ; de ce fait et de (ii) résulte la convergence faible dans L^2 , uniformément en x , des $U^h(x, \cdot)$ vers V ; mais alors,

$$\begin{aligned} E |U^h(X^h, \cdot) - V|^2 &= E |U^h(X^h, \cdot)|^2 + E |V|^2 - 2 E(V U^h(X^h, \cdot)) \\ &= \int E |U^h(x, \cdot)|^2 \phi^h(dx) + E |V|^2 - 2 \int E(V U^h(x, \cdot)) \phi^h(dx) \end{aligned}$$

et

$$\begin{aligned} \left| \int E |U^h(x, \cdot)|^2 \phi^h(dx) - E |V|^2 \right| &\leq \int |E |U^h(x, \cdot)|^2 - E |V|^2| \phi^h(dx) \\ &\leq \int \varepsilon \phi^h(dx) \text{ pour } h \text{ assez petit, indépendamment de } x. \\ &\leq \varepsilon \text{ (pour } \varepsilon > 0 \text{ quelconque)}. \end{aligned}$$

De même

$$\left| \int E(V(\cdot) U^h(x, \cdot)) \phi^h(dx) - E |V|^2 \right| \leq \int |E V(\cdot) U^h(x, \cdot) - E |V|^2| \phi^h(dx)$$

qui par convergence faible et l'inégalité de Cauchy-Schwarz est majoré par ε arbitrairement petit pour h suffisamment petit indépendamment de x . \square

Ce qui achève la démonstration du Lemme 3. \square

Etape 3 : Comparaison de W_n^h et Z_n

Considérons à présent la chaîne $(Z_n; n \geq 0)$ définie par

$$(3.2.16) \quad Z_{n+1} = r Z_n + s \varepsilon_{n+1}(e)$$

où, rappelons-le, $r = \lim_{h \rightarrow 0} r_h$ et $s = \lim_{h \rightarrow 0} s_h(0)$ avec :

$$(3.2.17) \quad \text{loi}(Z_0) = \text{loi stationnaire } \lambda \text{ de la chaîne } Z_n.$$

Lemme 5 :

Soit $u'_n = E^{1/(1+\beta)} |Z_n^h - Z_n|^{1+\beta}$ où β est un réel strictement compris entre 0 et 1.

En reprenant les notations précédentes, on a l'inégalité (3.2.19) :

$$\begin{aligned}
 E^{1/(1+\beta)} |W_n^h - Z_n|^{1+\beta} &\leq ||r||^n E^{1/(1+\beta)} |Z_0|^{1+\beta} + ||r||^n h^{-\frac{\gamma}{2}} v(1+\beta, h) \\
 &+ \text{Cste } ||r_n - r|| h^{-\frac{\gamma}{2}} v(1+\beta, h) + \text{Cste } ||s_0^h - s|| \\
 &+ \text{Cste } [C h^{\alpha/2(1+\alpha)} (1-R(h))^{-1} + h^{-1/2(1+\alpha)} v(\gamma, h)]^{1+\alpha} \\
 &+ \text{Cste } [C h^{\frac{\gamma}{2}(1-R(h))^{-1}} + v(\delta, h)]^\alpha \\
 &+ a(h)
 \end{aligned}$$

avec $\lim_{h \rightarrow 0} a(h) = 0$.

Démonstration :

On a :

$$Z_{n+1}^h - Z_{n+1} = r(Z_n^h - Z_n) + (r_n - r) Z_n^h + (s_n(0) - s) \varepsilon_{n+1}(e)$$

D'après l'inégalité de Minkowski et l'annexe 1, on a :

$$u_{n+1}' \leq ||r|| u_n' + ||r_n - r|| E^{1/(1+\beta)} |Z_n^h|^{1+\beta} + ||s_n(0) - s||$$

En particulier,

$$u_1' \leq ||r|| u_0' + ||r_n - r|| E^{1/(1+\beta)} |W_0^h|^{1+\beta} + ||s_n(0) - s||$$

et on a,

$$u_0' = E^{1/(1+\beta)} |Z_0^h - Z_n|^{1+\beta} \leq E^{1/(1+\beta)} |Z_0^h|^{1+\beta} + E^{1/(1+\beta)} |Z_0|^{1+\beta}$$

$$\leq h^{-\frac{\gamma}{2}} v(1+\beta; h) + \text{Cste d'après l'annexe 1.}$$

Il s'en suit aisément que :

$$\begin{aligned}
 u'_n &\leq ||r||^n u'_0 + \frac{1}{(1-||r||)} [||r_h - r|| h^{-\frac{\gamma}{2}} v(1+\beta, h) + ||s_n(0) - s||] \\
 (3.2.18) \quad u'_n &\leq ||r||^n E^{1/(1+\beta)} |Z_0|^{1+\beta} + ||r||^n h^{-\frac{\gamma}{2}} v(1+\beta, h) \\
 &\quad + \text{Cste} ||r_h - r|| h^{-\frac{\gamma}{2}} v(1+\beta, h) + \text{Cste} ||s_n(0) - s||
 \end{aligned}$$

Finalement la conjonction de (3.2.18) et de la relation (3.2.15) de l'étape 2 conduit à l'inégalité (3.2.19). \square

Etape 4 : Majoration des $E^{1/(1+\beta)} |V_n^{h1+\beta}|$ en régime stationnaire.

Lemme 6 :

En régime stationnaire, pour tout β , ($0 < \beta < 1$), il existe une constante telle que :

$$(3.2.20) \quad E^{1/(1+\beta)} |V_n^{h1+\beta}| \leq h^{\frac{\gamma}{2}} \text{Cste} \text{ où } V^h \text{ désigne une v.a. de loi } v^h \text{ (loi stationnaire de la chaîne } (W_n^h; n \geq 0)).$$

Démonstration :

Choisissons d'abord pour loi de V_0^h une mesure de Dirac. Cela va nous permettre de majorer, en régime stationnaire, les moments du type $E^{1/(1+\beta)} |V_n^{h1+\beta}|$.

Dans ce cas $v(\gamma, h) = 0$ pour tout γ ($1 < \gamma \leq 2$) et tout $h > 0$ l'inégalité (3.2.19) se simplifie en :

$$\begin{aligned}
 E^{1/(1+\beta)} |W_n^h - Z_n^h|^{1+\beta} &\leq ||r||^n \text{Cste} + \text{Cste} ||s_0^h - s|| + \text{Cste} h^{\alpha/2} (1-R(h))^{-(1+\alpha)} \\
 &\quad + \text{Cste} h^{\alpha/2} (1-R(h))^{-\alpha} + a(h)
 \end{aligned}$$

D'après l'hypothèse (3.2.1) du théorème, le membre de droite tend vers $\text{Cste} ||r||^n$ lorsque h tend vers 0, à n fixé.

D'après l'inégalité triangulaire, on a :

$$E^{1/(1+\beta)} |W_n^h|^{1+\beta} \leq E^{1/(1+\beta)} |W_n^h - Z_n^h|^{1+\beta} + E^{1/(1+\beta)} |Z_n^h|^{1+\beta}$$

Donc, d'après ce qui précède, $E^{1/(1+\beta)} |W_n^h|^{1+\beta}$ reste borné pour h assez petit. Donc, en régime stationnaire, on a bien

$$E^{1/(1+\beta)} |V^h|^{1+\beta} \leq h^{\frac{\gamma}{2}} \text{Cste où } V^h \text{ est une v.a. de loi } \nu^h.$$

Etape 5 :

Choisissons à présent pour loi de V_0^h la loi stationnaire μ^h . D'après (3.2.20), on a donc :

$$(3.2.21) \quad v(\gamma, h) < h^{\frac{\gamma}{2}} \text{Cste pour tout } \gamma \text{ de }]1, 2].$$

L'inégalité (3.2.19) devient donc :

$$\begin{aligned} E^{1/(1+\beta)} |W_n^h - Z_n^h|^{1+\beta} &\leq \text{Cste } ||r||^n \\ &+ \text{Cste } ||r||^n + \text{Cste } [C h^{\alpha/2(1+\alpha)} (1-R(h))^{-1} + h^{\alpha/2(1+\alpha)} \text{Cste}]^{1+\alpha} \\ &+ \text{Cste } ||r_h - r|| + \text{Cste } ||s_h(0) - s|| \\ &+ \text{Cste } [C h^{\frac{\gamma}{2}} (1-R(h))^{-1} + h^{\frac{\gamma}{2}} \text{Cste}]^\alpha + a(h) \end{aligned}$$

D'après l'hypothèse (3.2.1), il s'en suit que

$$(3.2.22) \quad E^{1/(1+\beta)} |W_n^h - Z_n^h|^{1+\beta} \leq \text{Cste } ||r||^n + c(h) \text{ avec } \lim_{h \rightarrow 0} c(h) = 0$$

Choisissons alors $n = n(h)$ qui tend vers l'infini quand h tend vers 0 ; de (3.2.1) et (3.2.22) on déduit que :

$$(3.2.23) \quad E^{1/(1+\beta)} |W_{n(h)}^h - Z_{n(h)}^h|^{1+\beta} \text{ tend vers 0 quand } h \text{ tend vers 0.}$$

Etape 6 :

On sait que la loi de $W_{n(h)}^h$ tend vers ν^h quand h tend vers 0 et que la

loi de $Z_{n(h)}$ tend vers λ quand h tend vers 0. De (3.2.23) il ressort que $W_{n(h)}^h - Z_{n(h)}$ tend vers 0 en probabilité quand h tend vers zéro.

D'après un lemme classique (cf [Fel 71] lemme 2 page 254), il s'en suit que :

$$\lim_{h \rightarrow 0} \text{étroite } v^h = \lambda. \text{ Ceci achève la démonstration du théorème. } \square$$

Remarques :

Le théorème reste vrai pour $s = 0$; dans ce cas λ est la mesure de Dirac δ_0 (notations centrées). Il est vrai aussi si s est une matrice non inversible, pourvu que la chaîne $(Z_n; n \geq 0)$ définie par (3.1.16) soit ergodique.

3.3 Loi limite de la loi stationnaire de (SEM)_h

Il nous faut maintenant montrer que les deux chaînes de Markov $(SEM)_h$ et $(V)_h$ sont asymptotiquement équivalentes.

Ce point fait l'objet du théorème 2 ci-dessous.

Nous allons considérer ici une loi de relance, lorsque X_n^h sort de G_n , pour l'algorithme SEM induite par la modélisation de la chaîne $(SEM)_h : (X_n^h; n \geq 0)$ par la chaîne $(V)_h : (V_n^h; n \geq 0)$.

Supposons que $X_n^h = V_n^h$.

Soit $X_{n+\frac{1}{2}}^h = T_h(X_n^h) + h^{\frac{1}{2}} s_h(X_n^h) \eta_{n+1}^h(X_n^h; e)$

(3.3.1) Si $X_{n+\frac{1}{2}}^h \in G_h$, posons $X_{n+1}^h = X_{n+\frac{1}{2}}^h$

Sinon, posons $X_{n+1}^h = V_{n+1+j}^h$ où j est le plus petit entier tel que $V_{n+1+j}^h \in G_h$.

Et ainsi de suite...

Cet entier j est fini presque sûrement car les réalisations successives $(V_{n+1}^h, V_{n+2}^h, \dots, V_{n+1+j}^h)$ sont indépendantes et leurs lois chargent G_h d'une masse non nulle et identique.

Remarques :

Dans le cas où $v_h = 0$, on a toujours $j = 1$ et la loi de V_{n+1}^h est la masse de Dirac au point t_h .

Cette procédure de relance pour la chaîne $(SEM)_h$ est particulière. Actuellement, nous ne savons pas prouver le théorème 2 pour une loi de relance γ_h quelconque. Cette restriction est essentiellement technique.

Théorème 2 :

Soit ϕ^h la loi limite de la chaîne $(SEM)_h$ ainsi définie et μ^h la loi limite de la chaîne $(V)_h$.

Sous les hypothèses du théorème 1 du paragraphe 3.2, on a :

(3.3.2) La restriction de μ^h à G_h est plus petite que ϕ^h (le support de ϕ^h est inclus dans G_h par définition).

En conséquence :

(3.3.3) La variation totale de la mesure signée $(\mu^h - \phi^h)$ tend vers 0 quand h tend vers 0 d'après le théorème 1.

Et par suite,

(3.3.4) $\lim_{h \rightarrow 0} \text{étroite } \psi^h = \lambda$

ψ^h étant la loi limite de la chaîne $(SEM)_h$ normalisée $(Y_n^h; n \geq 0)$.

Démonstration :

Le seul point difficile est d'établir (3.3.2). L'idée est de suivre une

trajectoire de $(X_n^h; n \geq 0)$ et la trajectoire correspondante de la chaîne $(V_n^h; n \geq 0)$: tant que V_n^h reste dans G_h , on a $V_n^h = X_n^h$; après l'instant de première sortie de G_h , on a $V_{n+1+j}^h = X_{n+1}^h$ (j étant l'entier défini dans (3.3.1)) et ainsi de suite.

Ainsi les V_n^h passent par les mêmes points de G_h que les X_n^h , mais avec un retard croissant. Plus précisément, par ergodicité, on a, si A désigne un borélien de R^D contenu dans G_h .

$$(3.3.5) \quad \mu^h(A) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n 1|_A(V_k^h)$$

$$(3.3.6) \quad \phi^h(A) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n 1|_A(X_k^h)$$

$1|_A$ étant la fonction indicatrice de A .

Si l'on note o_i les instants de sorties successifs de G_h pour la suite $(V_n^h; n \geq 0)$ et $o_i + j_i$ les instants de rentrée correspondants, on trouve, pour $i \geq 1$, si $V_0^h = X_0^h$:

$$(3.3.7) \quad \begin{aligned} & - \text{pour } n \leq o_1 - 1 : V_n^h = X_n^h \\ & - \text{pour } o_1 \leq n \leq o_1 + j_1 - 1 : V_n^h \notin G_h \implies V_n^h \notin A \text{ (} j_1 \geq 1 \text{)} \\ & - \text{pour } o_1 + j_1 \leq n \leq o_2 - 1 : V_n^h = X_{n-j_1}^h \\ & - \text{pour } o_i \leq n \leq o_i + j_i - 1 : V_n^h \notin G_h \implies V_n^h \notin A \text{ (} j_i \geq 1 \text{)} \\ & - \text{pour } o_i + j_i \leq n \leq o_{i+1} - 1 : V_n^h = X_{(n - \sum_{\ell=1}^i j_\ell)}^h \end{aligned}$$

Ainsi, pour tout k , ou bien $1|_A(V_k^h) = 0$, ou bien $\exists \ell \leq k$ tel que $1|_A(V_k^h) = 1|_A(X_\ell^h) = 1$.

D'où par passage à la limite $\mu^h(A) \leq \phi^h(A)$.

Nous pouvons maintenant conclure.

La partie positive de $\mu^h - \phi^h$ a pour support G_h^c où elle vaut $\mu^h(G_h^c)$; la partie négative a pour support G_h .

La variation totale de cette mesure est

$$\mu^h(G_h^c) + (\phi^h - \mu^h)(G_h)$$

et
$$\mu^h(G_h^c) = 1 - \mu^h(G_h)$$

$$\mu^h(G_h) + \phi^h(G_h) - \mu^h(G_h) = 1 \text{ car } \phi^h \text{ est à support inclus dans } G_h.$$

D'où
$$\mu^h(G_h^c) = (\phi^h - \mu^h)(G_h).$$

La variation totale de $\mu^h - \phi^h$ vaut donc $2\mu^h(G_h^c)$.

De plus $\lim_{h \rightarrow 0} \mu^h(G_h^c) = 0$ d'après le théorème 1 d'où (3.3.3).

Maintenant, les relations (3.3.2) et (3.3.3) s'appliquent aux mesures ψ^h et ν^h des chaînes normalisées.

Or d'après le théorème 1, $\lim_{h \rightarrow 0} \nu^h = \lambda$ d'où (3.3.4).

Remarque :

Dans les conditions du théorème 2, pour tout borélien A de R^D contenu dans G_h et tel que $\lambda(\text{FR}(A)) = 0$, on a :

$$(3.3.8) \quad |\psi^h(A) - \lambda(A)| \leq |\nu^h(A) - \lambda(A)| + \mu^h(G_h^c)$$

qui tend vers 0 quand h tend vers 0.

En effet :

$$0 \leq \phi^h(A) - \mu^h(A) \leq \mu^h(G_h^c)$$

$$\begin{aligned}
|\psi^h(A) - \lambda(A)| &\leq |\psi^h(A) - \nu^h(A)| + |\nu^h(A) - \lambda(A)| \\
&\leq \nu^h(G_h^c) + |\nu^h(A) - \lambda(A)|
\end{aligned}$$

d'où le résultat d'après le théorème du portemanteau de Billingsley (cf [Bi 68] page 12).

Il ressort des théorèmes 1 et 2 des paragraphes 3.2 et 3.3 que lorsque la taille de l'échantillon tend vers l'infini, à l'état stationnaire, la suite des estimés par l'algorithme SEM des paramètres d'un mélange de densités appartenant à la famille exponentielle, le nombre de composants étant connu, peut être considérée comme une suite de réalisations indépendantes d'une loi normale dont les caractéristiques sont :

- Le vecteur moyen est égal à la valeur de l'estimateur du maximum de vraisemblance, convergent quand N tend vers l'infini vers la vraie valeur des paramètres du mélange.

- La matrice variance est de forme $\frac{\Gamma}{\sqrt{N}}$.

La matrice Γ peut s'exprimer, de manière théorique, en fonction des caractéristiques du mélange.

3.4 Conjugaison par difféomorphisme

Soit g un difféomorphisme de classe C^2 de \mathbb{R}^p dans \mathbb{R}^p , tel que $g(0) = 0$, dont les différentielles première et seconde sont bornées sur \mathbb{R}^p ainsi que les différentielles première et seconde de g^{-1} .

On s'intéresse à la chaîne transformée $(V-g)_h$ définie par :

$$(3.4.1) \quad G_{n+1}^h = U_h(G_n^h) + h^{\frac{1}{2}} t_h(G_n^h) \zeta_{n+1}^h(G_n^h, e) \text{ où :}$$

$$U_h = g \circ T_h \circ g^{-1}$$

$$t_h(x) = D_g(T_h \circ g^{-1}(x)) s_h(g^{-1}(x))$$

$$\zeta_n^h(x, e) = \eta_n^h(g^{-1}(x), e)$$

Nous allons montrer que la chaîne normalisée $(H_n^h; n \geq 0)$ où $H_n^h = h^{-\frac{1}{2}} G_n^h$ vérifie un théorème de loi limite analogue au théorème 1. Nous supposons, comme en (3.1.11) que la chaîne $(V-g)_h$ est ergodique, de loi stationnaire τ^h ; nous noterons θ^h la loi stationnaire de la chaîne normalisée $(H_n^h; n \geq 0)$. Le ressort de cette preuve est la comparaison de la chaîne $(H_n^h; n \geq 0)$ à la chaîne déduite du processus image par g de $(V)_h$, normalisée : on notera

$$(3.4.2) \quad \begin{aligned} J_n^h &= g(V_n^h) \\ K_n^h &= h^{-\frac{1}{2}} g(V_n^h) \end{aligned}$$

On a successivement

$$\begin{aligned} g(V_{n+1}^h) &= g[T_h(V_n^h) + h^{\frac{1}{2}} s_h(V_n^h) \eta_{n+1}^h(V_n^h; e)] \\ |g(V_{n+1}^h) - U_h(g(V_n^h)) - h^{\frac{1}{2}} D_g(T_h \circ g^{-1}(g(V_n^h))) s_h(g^{-1}(g(V_n^h))) \eta_{n+1}^h(g^{-1}(g(V_n^h)); e)| \\ &\leq h \text{ Cste} | \eta_{n+1}^h(g^{-1}(g(V_n^h)); e) |^2 \end{aligned}$$

en raison de l'inégalité de Taylor, de l'hypothèse que $D^2 g$ est bornée sur R^p et (H11). Cela s'écrit :

$$(3.4.3) \quad \begin{aligned} |J_{n+1}^h - U_h(J_n^h) - h^{\frac{1}{2}} D_g(T_h \circ g^{-1}(J_n^h)) s_h(g^{-1}(J_n^h)) \zeta_{n+1}^h(J_n^h; e)| \\ \leq h \text{ Cste} | \zeta_{n+1}^h(J_n^h; e) |^2 \end{aligned}$$

D'après (3.1.8) et en utilisant les notations ci-dessus

$$E | J_{n+1}^h - U_h(J_n^h) - h^{\frac{1}{2}} t_h(J_n^h) \zeta_{n+1}^h(J_n^h; e) | < \text{Cste} h.$$

En divisant par $h^{\frac{1}{2}}$, on obtient

$$(3.4.4) \quad \begin{aligned} E | K_{n+1}^h - h^{\frac{1}{2}} U_h(h^{\frac{1}{2}} K_n^h) - [t_h(h^{\frac{1}{2}} K_n^h) - t_h(0)] \zeta_{n+1}^h(h^{\frac{1}{2}} K_n^h; e) + \\ + t_h(0) \zeta_{n+1}^h(h^{\frac{1}{2}} K_n^h; e) | \\ < \text{Cste} h^{\frac{1}{2}} \end{aligned}$$

Soit $r'_h = D U_h(0) = D_g(0) D_{T_h}(0) D_g^{-1}(0) = D_g(0) r_h D_g^{-1}(0)$

De manière analogue à (3.2.6), on peut montrer que :

$$|h^{\frac{1}{2}} U_h(h^{\frac{1}{2}} K_n^h) - r'_h(y)| \leq \text{Cste } h^{\frac{1}{2}} |y|^2$$

D'où

$$(3.4.5) \quad E |h^{-\frac{1}{2}} U_h(h^{\frac{1}{2}} K_n^h) - r'_h(K_n^h)|^2 \leq \text{Cste } h^{\frac{1}{2}} E |K_n^h|^2$$

Or la suite des $E |K_n^h|^2$ reste bornée. En effet :

$$E |K_n^h|^2 = h^{-1} E |J_n^h|^2 = h^{-1} E |g(V_n^h)|^2 = h^{-1} E |g(V_n^h) - g(0)|^2$$

d'où

$$(3.4.6) \quad E |K_n^h|^2 \leq \text{Cste } h^{-1} E |V_n^h|^2 \text{ car } g \text{ est Lipschitzienne puisque } D_g \text{ est borné.}$$

D'après (3.2.20), il s'en suit que

$$(3.4.7) \quad E |K_n^h|^2 \leq \text{Cste}$$

On a ainsi prouvé que

$$(3.4.8) \quad E |h^{-\frac{1}{2}} U_h(h^{\frac{1}{2}} K_n^h) - r'_h(K_n^h)|^2 \leq \text{Cste } h^{\frac{1}{2}}$$

De même, de manière analogue à (3.2.7) puis du passage de (3.4.5) à (3.4.8) on montre que

$$(3.4.9) \quad E |t_h(h^{\frac{1}{2}} K_n^h) - t_h(0)| \leq \text{Cste } h^{\frac{1}{2}}$$

Finalement, par application de l'inégalité triangulaire à (3.4.4) et par (3.4.8) et (3.4.9), il vient :

$$(3.4.10) \quad E |K_{n+1}^h - r'_h(K_n^h) - t_h(0) \zeta_{n+1}^h(h^{\frac{1}{2}} K_n^h; e)| \leq \text{Cste } h^{\frac{1}{2}}$$

Par comparaison avec $(V-g)_n$, si l'on se place sur l'espace $(W, B(W), m)$, comme en 3.2 étape 2, et si l'on suppose que $G_0^h = J_0^h$:

$$\begin{aligned} E | K_{n+1}^h - H_{n+1}^h | &\leq q E | K_n^h - H_n^h | + Cste h^{\frac{1}{2}} \\ &+ Cste E | \zeta_{n+1}^h (h^{\frac{1}{2}} K_n^h; w) - \varepsilon_{n+1}(w) | \\ &+ Cste E | \zeta_{n+1}^h (h^{\frac{1}{2}} H_n^h; w) - \varepsilon_{n+1}(w) | \end{aligned}$$

Raisonnant comme pour (3.2.14) on obtient

$$E | K_{n+1}^h - H_{n+1}^h | \leq q E | K_n^h - H_n^h | + Cste h^{\frac{1}{2}} + a(h)$$

avec $\lim_{h \rightarrow 0} a(h) = 0$ (et, rappelons-le, $0 \leq q < 1$, voir (H9)).

Il en résulte

$$(\forall n \geq 0) (\forall h > 0) E | K_n^h - H_n^h | \leq (Cste h^{\frac{1}{2}} + a(h)) Cste$$

Ainsi on pourra conclure comme dans la preuve du théorème 1.

Nous sommes donc en mesure d'énoncer le théorème suivant.

Théorème 3

Soit g un difféomorphisme C^2 de \mathbb{R}^D dans \mathbb{R}^D tel que $g(0) = 0$, dont les différentielles première et seconde sont bornées sur \mathbb{R}^D ainsi que celles de g^{-1} . Sous les hypothèses (H1) à (H17) et si la chaîne $(G_n^h; n \geq 0)$ définie par

$$(V-g)_n \quad G_{n+1}^h = g \circ T_n \circ g^{-1} (G_n^h) + h^{\frac{1}{2}} s_n (G_n^h) \eta_{n+1}^h (G_n^h, e)$$

est ergodique, alors la loi stationnaire θ_n de la chaîne $(H_n^h; n \geq 0)$ définie par $H_n^h = h^{-\frac{1}{2}} G_n^h$ converge, lorsque $h \rightarrow 0$, vers la loi gaussienne de centre 0, et de matrice variance $D_g(0) \left[\sum_{\ell=0}^{\ell} r^\ell s s'(r')^\ell \right] D_g'(0)$. (le signe ' désignant la transposée).

Démonstration

D'après ce qui précède, il suffit de montrer que la chaîne $(K_n^h; n > 0)$ a même loi limite que la chaîne $(D_g(0) W_n^h; n > 0)$.

$$\text{En effet, } K_n^h = h^{-\frac{1}{2}} g(h^{\frac{1}{2}} W_n^h)$$

$$\text{d'où } K_n^h = h^{-\frac{1}{2}} g(0) + h^{-\frac{1}{2}} h^{\frac{1}{2}} D_g(0) W_n^h + R_n^h \quad \text{avec } g(0) = 0$$

$$\text{où } E | R_n^h | \leq \text{Cste } h^{\frac{1}{2}} \text{ et donc}$$

$$E | K_n^h - D_g(0) W_n^h | \leq \text{Cste } h^{\frac{1}{2}}. \quad \square$$

Maintenant pour déduire du théorème 3 un résultat analogue au théorème 2, on construit la chaîne $(SEM-g)_h$ en la déduisant de $(V-g)_h$ au moyen d'une loi de relance définie comme en 3.3. Un raisonnement analogue permet alors d'obtenir :

Théorème 4

Avec les mêmes notations et sous les mêmes hypothèses que ci-dessus, la loi stationnaire de la chaîne $(h^{-\frac{1}{2}} L_n^h; n \geq 0)$ - où $(L_n^h; n \geq 0)$ est la chaîne $(SEM-g)_h$ déduite de $(V-g)_h$ comme en 3.3 - converge étroitement, lorsque $h \rightarrow 0$, vers la loi gaussienne de centre 0 et de matrice de variance

$$D_g(0) \left[\sum_{l=0}^{\infty} r^l s s'(r^l) \right] D_g'(0).$$

3.5 Portée des résultats obtenus

3.5.1 Sur l'algorithme EM

Sur le plan mathématique, l'algorithme EM apparaît comme une famille de systèmes dynamiques à temps discrets $(T_h; h > 0)$ où

$$T_h : \text{int}(G) \rightarrow \text{int}(G) \quad (\text{int} : \text{intérieur})$$

G étant une sous-variété différentielle bornée à bord de l'espace \mathbb{R}^P . Le paramètre $h = 1/N$ désigne l'ensemble des échantillons de taille N .

Le bord de la variété G consiste en les paramètres $(p_1, \dots, p_K, a_1, \dots, a_K)$ pour lesquels un au moins des paramètres de proportion p_k est nul : il s'agit donc des mélanges dont le nombre de composants est inférieur à K .

Notons ici, $p = (p_1, \dots, p_K)$ et $a = (a_1, \dots, a_K)$.

Une difficulté de l'étude du comportement des systèmes dynamiques T_h provient de ce que l'écriture explicite de l'expression de $T_h(p, a)$ en fonction de (p, a) est d'une lourdeur telle qu'aucune propriété précise de T_h ne peut en être déduite.

Suivant les travaux de [ReWa 84], [Wu 83] il nous paraît raisonnable de schématiser les propriétés des systèmes dynamiques T_h ainsi :

(i) la log-vraisemblance du mélange, que nous avons noté L , (voir 2.1) est une fonction de Liapunov :

$$(\forall x) \text{ si } x \text{ n'est pas point fixe de } T_h \text{ alors } L \circ T_h(x) > L(x)$$

(ii) chaque fonction T_h admet un unique point fixe attractif, que nous avons supposé être le point 0 de \mathbb{R}^P (notations centrées, voir 2.6.1),

(iii) chaque fonction T_h n'admet qu'un nombre fini de points fixes,

(iv) toute trajectoire converge vers l'un de ces points fixes.

Les limitations fondamentales de l'algorithme EM (voir 2.5, et pour plus de détails [CeDi 84]) se traduisent ici par :

- puisque T_h envoie $\text{int}(G)$ dans $\text{int}(G)$ le nombre de composants du mélange étudié reste constant d'itération en itération : le nombre exact de composants du mélange doit être connu avant la mise en oeuvre de l'algorithme.

- L'existence de points fixes non attractifs de T_h vers lesquels certaines trajectoires convergent (vérifiée expérimentalement) ne permet pas de garantir la validité du résultat obtenu. De plus, la convergence vers de tels points fixes peut se faire avec une lenteur rédhibitoire (voir [CeDi 84], [ReWa 84]).

3.5.2 Sur l'algorithme SEM

Sur le plan mathématique, l'algorithme SEM, qui répond à ces limitations, se présente comme une perturbation aléatoire, dont la variance est de l'ordre de h (donc tend vers zéro quand $h \rightarrow 0$), de la famille de système dynamiques $(T_h; h > 0)$. Les itérés X_n^h du système perturbé forment une chaîne de Markov ergodique, que nous avons notée $(SEM)_h$ au paragraphe 3.1. Ces perturbations s'accompagnent d'un mécanisme de relance aléatoire lorsque X_n^h s'approche trop du bord de G , i.e. lorsqu'une des classes construites par l'algorithme SEM tend à se vider. Plus précisément la dimension d des vecteurs-échantillon étant fixée, nous définissons, pour chaque taille N de l'échantillon, un seuil $0 < c(N) < 1$, qui tend vers 0 quand $N \rightarrow \infty$; lorsqu'une des proportions devient inférieure à ce seuil, nous considérons que le composant correspondant a disparu; nous enregistrons l'évènement puis relançons l'algorithme en tirant la v.a. X_{n+1}^h , indépendamment du passé, selon une loi de probabilité γ_h , fixée à l'avance, à support contenu dans l'ensemble G_h des paramètres (p, a) de G tels que toutes les proportions soient plus grandes que $c(N)$: les parties G_h de G croissent vers G lorsque h décroît vers 0.

Ainsi que nous l'avons expliqué aux paragraphes 2.6 et 3.1, il nous paraît raisonnable de schématiser en les résumant les propriétés des systèmes dynamiques perturbés $(SEM)_h$ en adjoignant aux hypothèses (i), (ii), (iii), (iv) (portant sur T_h) les hypothèses suivantes :

(v) les perturbations ont la forme $h^{1/2} s_h(x) \eta_n^h(x; e)$ (où $x \in G_h$) : voir les hypothèses précises et les notations en 3.1,

(vi) les matrices $s_h(x)$, ainsi que leur limite $s(0)$ sont régulières

(vii) la famille des bruits $(\eta_n^h(x, e); n \geq 1)$ converge uniformément en $x \in G_h$ vers le bruit blanc gaussien standard

(viii) les matrices $s_n(x)$ tendent à dégénérer au bord de G lorsque $h \rightarrow 0$.

3.5.3 Les problèmes posés

Les problèmes mathématiques posés se résument ainsi :

Problème I : Prouver que la loi stationnaire de la chaîne normalisée $(X_n^h h^{-1/2} = Y_n^h; n \geq 0)$ converge étroitement lorsque h tend vers 0 vers une loi gaussienne centrée de variance non dégénérée.

Problème II : Déterminer la forme asymptotique, quand h tend vers 0, des lois des temps entre deux atteintes successives du bord de G_h par les v.a. X_n^h .

3.5.4 Le problème traité

Dans ce travail, nous n'avons étudié que le problème I. Nous ne l'avons pas résolu dans sa généralité. Nous avons obtenu (théorème 4) le résultat escompté :

- En remplaçant les hypothèses (les propriétés ?) (iii) et (iv) par les propriétés (iii bis) et (iv bis) :

(iii bis) chaque fonction T_h n'admet qu'un point fixe : l'unique point fixe attractif 0

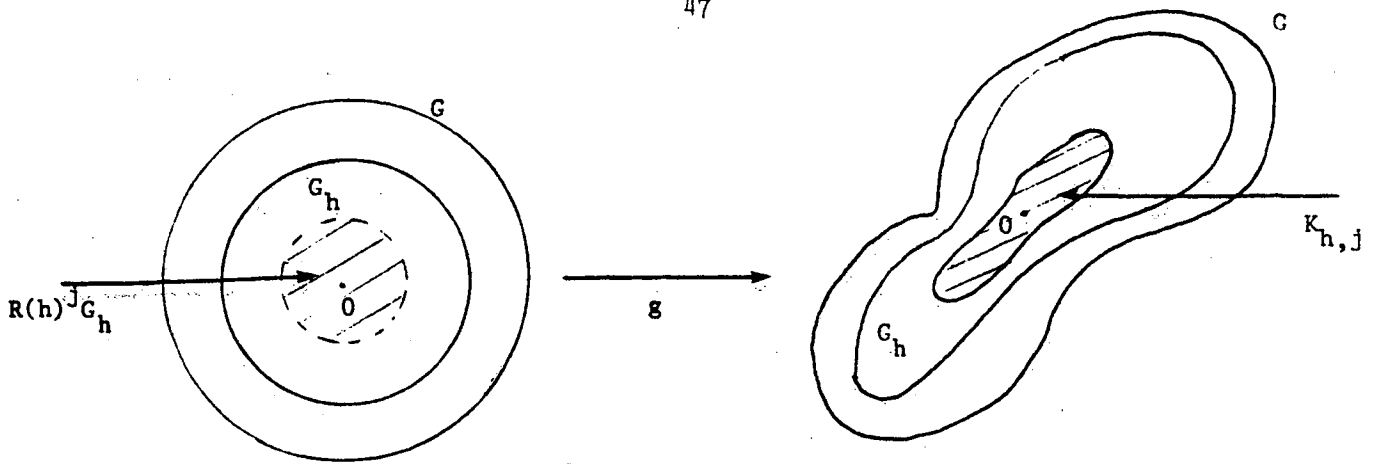
(iv bis) toute trajectoire converge vers ce point fixe.

- En adjoignant une propriété du type :

Pour chaque $h > 0$, il existe une suite exhaustive de compacts emboîtés de G_h , d'intersection $\{0\}$, que nous noterons $(K_{h,j}; j \geq 0)$, telle que

$$(\forall j \geq 0) \quad T_h(K_{h,j}) \subset K_{h,j+1}$$

les $K_{h,j}$ sont les images par le difféomorphisme g des ensembles $R(h)^j$. G_h



Ainsi l'intérêt des théorèmes démontrés en 3.4 réside dans le fait qu'il n'est pas nécessaire de supposer que $|T_h(x)| < |x|$ pour tout x .

- En choisissant pour lois de relance γ_h des lois appartenant à une classe particulière contenant les mesures de Dirac (voir paragraphe 3.2 étape 4). Cette dernière contrainte ne nous semble pas essentielle et devrait pouvoir être levée assez facilement.

3.5.5 Sur le seuil

En ce qui concerne le seuil $c(N) = c(h)$, la condition imposée (cf théorèmes 1 et 2) $\lim_{h \rightarrow 0} h^{1/4} (1 - R(h))^{-1} = 0$ signifie ceci : si les T_h admettent au voisinage du bord de G des différentielles uniformément bornées, alors $R(h)$ et $c(h)$ sont liés par une relation de la forme : $c(h)$ est proportionnelle à $1 - R(h)$ lorsque $h \rightarrow 0$. La condition imposée s'écrit alors

$$\lim_{h \rightarrow 0} h^{+1/4} c(h)^{-1} = 0.$$

On doit donc prendre $c(N)$ vérifiant :

$c(N) \rightarrow 0$ et $c(N) N^{1/4} \rightarrow \infty$ quand N tend vers l'infini.

Ceci suggère de choisir $c(N)$ nettement plus grand que $(d+1)/N$ dans le cas de grands échantillons.

4 - COMPORTEMENT ASYMPTOTIQUE D'UNE VERSION SEQUENTIELLE DE L'ALGORITHME SEM

Le principe d'une version séquentielle de l'algorithme SEM consiste à considérer que l'échantillon est construit par étapes successives. A chaque étape n , on observe le $n^{\text{ième}}$ point de l'échantillon x_n et l'on modifie les estimations des paramètres du mélange compte-tenu de ce nouveau point.

Deux versions séquentielles sont possibles (cf [CeDi 84]). La version "stricte" consiste à ne prendre en compte que le dernier point observé pour modifier les estimations. Au contraire, la version "large" utilise tout l'échantillon déjà observé $((x_1, \dots, x_n)$ à l'étape n) pour modifier les estimations.

Dans le cas d'un mélange à deux composants où seul le paramètre de proportion était inconnu, nous avons étudié le comportement asymptotique de ces deux versions séquentielles de l'algorithme dans [CeDi 84].

L'étude dans le cas séquentiel strict s'appuie sur un article de Silverman [Si 80] ; elle utilise des théorèmes de convergence de martingales et est difficilement généralisable au cas multidimensionnel.

Par contre, l'étude dans le cas séquentiel large utilise les mêmes outils que ceux utilisés au paragraphe 3. Nous allons donc reprendre dans le cas général (multidimensionnel) l'étude du comportement asymptotique de la version séquentielle large de l'algorithme SEM en utilisant la même modélisation que précédemment.

4.1 Définitions, notations et hypothèses

Dans le cas étudié ici, par construction même de l'algorithme, les indices n (dénotant le numéro d'itération) et h sont liés par la relation $h = \frac{1}{n}$. Il ne convient donc pas de les distinguer dans nos notations.

A cette différence de notation près, les formules et les hypothèses introduites sont les mêmes que celle présentées au paragraphe 3.1 pour le cas non séquentiel.

Ainsi les deux chaînes de Markov à étudier se définissent ainsi :

- La chaîne $(SEM)_n = (X_n; n \geq 0)$:

$$\text{soit } X_{n+\frac{1}{2}} = T_n(X_n) + n^{-\frac{1}{2}} s_n \eta_{n+1}(X_n; e)$$

$$X_{n+1} = X_{n+\frac{1}{2}} \text{ si } X_{n+\frac{1}{2}} \in G_n$$

$X_{n+1} = \Gamma_{n+1}$ sinon, Γ_{n+1} étant une réalisation tirée indépendamment des v.a. X_0, \dots, X_n et de $\eta_{n+1}(X_n, \dots)$ selon une loi de probabilité γ_n à support dans le borélien $R(n) G_n$.

- La chaîne $(V)_n = (V_n; n \geq 0)$:

$$V_{n+1} = T_n(V_n) + n^{\frac{1}{2}} s_n(V_n) \eta_{n+1}(V_n; e)$$

$$\text{avec } T_n(x) = t_n \text{ si } x \notin G_n$$

$$\text{et } s_n(x) = v_n \text{ si } x \notin G_n.$$

Les chaînes auxiliaires utilisées dans la démonstration du théorème 5 ci-dessous (analogue du théorème 1) diffèrent légèrement de celles définies au paragraphe 1.

On introduit les trois chaînes de Markov suivantes :

- $(Z_n; n \geq 0)$ définie par :

$$Z_{n+1} = q_n r_n Z_n + q_n s_n(0) \varepsilon_{n+1}(e) \text{ et } Z_0 = W_0 = n^{\frac{1}{2}} V_0$$

où l'on a posé $q_n = (n+1)^{\frac{1}{2}} n^{-\frac{1}{2}}$ si $n > 0$ et $q_0 = 1$.

On supposera que :

(H'18) La chaîne $(Z_n; n \geq 0)$ est ergodique

- $(Z_n^*; n \geq 0)$ définie par :

$$Z_{n+1}^* = q_n r Z_n^* + q_n s \varepsilon_{n+1} \quad (e) \text{ et } Z_0^* = 0.$$

On supposera que :

(H'19) La chaîne $(Z_n^*; n \geq 0)$ est ergodique

- $(Z_n^{**}; n \geq 0)$ définie par :

$$Z_{n+1}^{**} = r Z_n^{**} + s \varepsilon_{n+1} \quad (e)$$

Et l'on suppose de même :

(H'20) La chaîne $(Z_n^{**}; n \geq 0)$ est ergodique. On notera λ sa mesure de probabilité stationnaire et l'on supposera loi $(Z_0^{**}) = \lambda$.

4.2 Loi limite de la loi stationnaire de $(W_n; n \geq 0)$

Théorème 5 :

Sans les hypothèses (H1) à (H17), (H'18) à (H'20), si l'on suppose en outre que :

(i) $\sum_{k \geq 0} R(k) = 0$

(ii) $\lim_{n \rightarrow \infty} n^{1/2} U_n^2 = 0$ où

$$U_n = \sum_{k=0}^{n-1} R(n-1) R(n-2) \dots R(n-k-1) \quad (n-k-1)^{-1/2}$$

avec la convention $R(0) = 1$

Alors on a :

$$\lim_{n \rightarrow \infty} \text{étroite } v_n = \lambda, \text{ d'où}$$

V_n tend vers zéro en probabilité quand n tend vers l'infini.

Démonstration :

On reprend point par point celle du théorème 1.

Etape 1 : Quelques inégalités.

Notons $u_n = E^{1/\gamma} |V_n|^\gamma$ pour $1 < \gamma < 2$ de sorte que $u_0 = v(\gamma)$;

on obtient (de manière analogue à (3.2.5)) :

$$(4.1.1) \quad u_{n+1} \leq R(n) u_n + c n^{-\frac{1}{2}} \text{ pour } n \geq 0 ; \text{ on en déduit aisément :}$$

$$(4.1.2) \quad u_n \leq R(0) R(1) \dots R(n-1) v(\gamma) + C U_n.$$

D'autre part, on obtient, de manière analogue à (3.2.6) et (3.2.7) :

Pour tout $x = n^{-\frac{1}{2}} y$ et tout $\alpha \in]0, 1]$

$$(4.1.3) \quad |n^{\frac{1}{2}} T_n(n^{-\frac{1}{2}} y) - r_n y| \leq \text{Cste } n^{-\alpha/2} |y|^{1+\alpha} \text{ et}$$

$$(4.1.4) \quad |s_n(n^{-\frac{1}{2}} y) - s_n(0)| \leq \text{Cste } n^{-\alpha/2} |y|^\alpha$$

Etape 2 : Comparaison de W_n et Z_n

De manière analogue à l'étape 2 de la démonstration du théorème 1, nous allons travailler à partir de v.a. $\hat{\eta}_n(x, w)$ et $\hat{\varepsilon}(w)$ définies, sur un espace probabilisé adéquat W , à partir de $\eta_n(x, e)$ et $\varepsilon_n(e)$ par le théorème de représentation de Skorohod. Ici aussi, nous confondrons dans l'écriture les v.a. définies sur W et les v.a. définies sur ξ .

Notons tout d'abord que puisque $\lim_{n \rightarrow \infty} q_n = 1$ et que $\limsup_{n \rightarrow \infty} ||r_n|| \leq q$ (d'après (H9)), il existe un entier n_0 et r_0 ($q < r_0 < 1$) tels que :

$$(4.1.5) \quad \forall n \geq n_0 \quad q_n ||r_n|| \leq r_0$$

Nous allons raisonner à partir de $n = n_0$.

Il est facile de voir que la chaîne normalisée ($W_n = n^{\frac{1}{2}} V_n$; $n \geq 0$) vérifie la relation :

$$(4.1.6) \quad W_{n+1} = q_n n^{\frac{1}{2}} T_n(n^{-\frac{1}{2}} W_n) + q_n s_n (n^{-\frac{1}{2}} W_n) \zeta_{n+1}$$

où rappelons-le $\zeta_{n+1} = \eta_{n+1} (n^{-\frac{1}{2}} W_n; e)$

Soit $t_n = E^{1/1+\beta} |W_n - Z_n|^{1+\beta}$ ($0 < \beta < 1$) et donc $t_0 = 0$.

Notons $a_n = E^{1/1+\beta} |\zeta_{n+1} - \zeta_{n+1}(e)|^{1+\beta}$ qui tend vers 0 quand n tend vers l'infini (cf 3.2.14).

Des relations (4.1.2), (4.1.3), (4.1.4), (4.1.5) et de l'inégalité de Minkowski, on tire que :

$$(4.1.7) \quad t_{n+1} \leq r_0 t_n + q_n [Cste (C_n^{1/2(1+\alpha)} U_n + n^{1/2(1+\alpha)} R(0)R(1)\dots R(n-1)v(\gamma))]^{1+\alpha} \\ + Cste [C U_n + R(0)R(1) \dots R(n-1) v(\delta)]^\alpha + Cste a_n$$

avec $\gamma = (1+\alpha)(1+\beta)$ et $\delta = \alpha(1+\beta)$.

Cette inégalité peut s'écrire sous la forme

$$(4.1.8) \quad t_{n+1} \leq r_0 t_n + q_n f_n \text{ avec } \lim_{n \rightarrow \infty} f_n = 0 \text{ par (i) et (ii).}$$

Comme $t_0 = 0$, on en déduit que :

$$(4.1.9) \quad t_n \leq \sum_{k=0}^{n-1} (r_0)^k q_{n-k-1} f_{n-k-1} \text{ d'où}$$

$$(4.1.10) \quad \lim_{n \rightarrow \infty} t_n = 0.$$

Etape 3 : Comparaison de Z_n et Z_n^*

Soit $u'_n = E^{1/1+\beta} |Z_n - Z_n^*|^{1+\beta}$ ($0 < \beta < 1$) et donc $u'_0 = v(1+\beta)$. D'après l'inégalité de Minkowski et (4.1.5), on a :

$$(4.1.11) \quad u'_{n+1} \leq r_0 u'_n + q_n [|r_n - r| E^{1/1+\beta} |Z_n^*|^{1+\beta} + |s_n(0) - s|]$$

Cette inégalité peut s'écrire sous la forme

$$(4.1.12) \quad u'_{n+1} \leq r_0 u'_n + q_n g_n \quad \text{avec} \quad \lim_{n \rightarrow \infty} g_n = 0, \quad \text{d'après (3.1.2) et (3.1.5)}$$

car $E^{1/1+\beta} |Z_n^*|^{1+\beta}$ reste borné pour n assez grand.

On en déduit :

$$(4.1.13) \quad u'_n \leq (r_0)^n v(1+\beta) + b_n \quad \text{avec} \quad b_n = \sum_{k=0}^{n-1} (r_0)^k q_{n-k-1} g_{n-k-1}$$

$$(4.1.14) \quad \text{d'où} \quad \lim_{n \rightarrow \infty} b_n = 0$$

Etape 4 : Comparaison de Z_n^* et Z_n^{**}

Soit $v_n = E^{1/1+\beta} |Z_n^* - Z_n^{**}|^{1+\beta}$ ($0 < \beta < 1$) et donc v_0 est une constante non nulle.

D'après l'inégalité de Minkowski, (4.1.5) et (H11), on a :

$$(4.1.15) \quad v_{n+1} \leq r_0 v_n + C |q_{n-1}| \quad \text{d'où}$$

$$(4.1.16) \quad v_n \leq \text{Cste} r_0^n + C \sum_{k=0}^{n-1} r_0^k |q_{n-k-1} - 1|$$

Etape 5 : Comparaison de W_n et Z_n^{**}

En regroupant les résultats des étapes 2, 3 et 4, il vient :

$$(4.1.17) \quad E^{1/1+\beta} |W_n - Z_n^{**}|^{1+\beta} \leq \text{Cste} r_0^n + t_n + b_n + \text{Cste} \sum_{k=0}^{n-1} r_0^{n-k-1} |q_k - 1|$$

qui tend vers zéro lorsque n tend vers l'infini d'après (4.1.5), (4.1.10) et (4.1.14).

Ceci achève la démonstration du théorème 3. \square

4.3 Loi limite de la loi stationnaire de $(SEM)_n$

Nous allons maintenant comparer les deux chaînes de Markov $(SEM)_n$ et $(V)_n$ comme il est fait dans le paragraphe 3.3.

Là encore, la démarche est tout à fait analogue au cas non séquentiel.

Nous considérons pour $(X_n; n \geq 0)$ la loi de relance suivante :

Supposons que $X_n = V_n$

Soit $X_{n+\frac{1}{2}} = T_n(X_n) + n^{-\frac{1}{2}} s_n(X_n) \eta_{n+1}(X_n; e)$

si $X_{n+\frac{1}{2}} \in G_n$ posons $X_{n+1} = X_{n+\frac{1}{2}}$

sinon posons $X_{n+1} = V_{n+1+j}$ où j est le plus petit entier tel que $V_{n+1+j} \in G_n$.

Théorème 6 :

Soit ϕ la loi limite de la chaîne $(SEM)_n$ ainsi définie et $\mu = \lambda$ la loi limite de la chaîne $(V)_n$.

Sous les hypothèses du théorème 3 du paragraphe 4.2, on a :

$$(4.3.1) \quad \phi = \lambda \text{ d'où}$$

(4.3.2) La limite en probabilité quand n tend vers l'infini de X_n est 0.

Démonstration :

La démonstration de ce théorème est identique à la démonstration du théorème 2 du paragraphe 3.3. Il suffit de la recopier en supprimant l'indice h .

Remarque :

Pour l'algorithme séquentiel strict, dans le cas unidimensionnel, on obtient la convergence p.s. du paramètre. Ici on a seulement une convergence en probabilité.

4.4 Conjugaison par difféomorphisme

On peut reprendre très facilement l'étude menée au paragraphe 3.4.

Sous les mêmes hypothèses et notations, les mêmes procédés conduisent à :

$$E |K_n - H_n| \leq Cste n^{-\frac{1}{2}} + a_n \text{ où } \lim_{n \rightarrow \infty} a_n = 0.$$

Par ailleurs on connaît le comportement de K_n :

$$K_n = n^{\frac{1}{2}} g(V_n) = n^{\frac{1}{2}} g(n^{-\frac{1}{2}} W_n) = n^{\frac{1}{2}} [g(0) + n^{-\frac{1}{2}} D_g(0) W_n + n^{-1} \dots]$$

avec $g(0) = 0$

$$(4.4.1) \quad \text{d'où } K_n = D_g(0) W_n$$

Par le théorème 5, on sait que la loi de W_n est une loi normale centrée de matrice variance non dégénérée.

D'après (4.4.1), c'est également le cas pour la loi de K_n .

En suivant toujours la même démarche qu'en 3.4, on obtient un théorème analogue au théorème 4.

En particulier :

- $(SEM-g)_n$, transformé de la chaîne $(SEM)_n$ par le difféomorphisme g a même loi limite que la chaîne $(V-g)_n$.

- La suite L_n définie par

ANNEXE 1

Proposition : Pour toute v.a. X de carré sommable définie sur un espace de probabilité quelconque l'expression $E^{1/\gamma} (|X|^\gamma)$ ($0 < \gamma \leq 2$) est une fonction croissante de γ .

Démonstration :

Soient deux réels γ et δ tels que $0 < \gamma < \delta \leq 2$. D'après l'inégalité de Hölder, $\forall p, q > 1$ tels que $\frac{1}{p} + \frac{1}{q} = 1$

$$E^{1/\gamma} (|X|^\gamma) = E^{1/\gamma} (1 \cdot |X|^\gamma) \leq E^{\frac{1}{p\gamma}} (1^p) E^{\frac{1}{q\gamma}} (|X|^{q\gamma})$$

Choisissons $q = \frac{\delta}{\gamma}$, il s'en suit que

$$E^{1/\gamma} (|X|^\gamma) \leq 1 \cdot E^{1/\delta} (|X|^\delta) = E^{1/\delta} (|X|^\delta).$$

ANNEXE 2

Soit b un nombre réel strictement compris entre 0 et 1. On peut choisir trois nombres réels a , p et q vérifiant les conditions suivantes :

- (1) $0 < a < 1$
- (2) $(1 + a)(1 + b) < 2$
- (3) $p, q > 1$ avec $\frac{1}{p} + \frac{1}{q} = 1$
- (4) $q(1 + b) < 2$
- (5) $1 < p a(1 + b) < 2$

Preuve :

Soit a quelconque vérifiant les conditions (1) et (2). Soit $q > 1$ vérifiant (4), alors $p = (1 - \frac{1}{q})^{-1} > \frac{2}{1-b}$.

Supposons maintenant que $q = \frac{2}{1+b}$; on a alors $p = \frac{2}{1-b}$.

De plus $pa(1+b) = \frac{2 a(1+b)}{(1-b)}$ et on en déduit que $pa(1+b) < 2$.

En effet, $pa(1+b) < 2 \iff a(1+b) < (1-b)$.

Or $(1+a)(1+b) < 2$, d'où $1 + b + a(1+b) < 2$ et donc $a(1+b) < 1-b$.

Reprenons l'hypothèse $q < \frac{2}{1+b}$, on a alors $pa(1+b) > \frac{2 a(1+b)}{(1-b)}$ quantité strictement inférieure à 2 d'après ce qui précède.

Par continuité, quand q est arbitrairement proche de $\frac{2}{1+b}$, $pa(1+b)$ est arbitrairement proche de $\frac{2 a(1+b)}{(1-b)}$.

Donc, pour avoir $pa(1+b) < 2$, il suffit de choisir q arbitrairement proche de $\frac{2}{1+b}$.

Il reste à vérifier que $pa(1+b) > 1$. Or $pa(1+b) > 2a \frac{(1+b)}{(1-b)}$ quantité plus grande ou égale à un si et seulement si $2a + (2a+1)b \geq 1$.

La condition $(1+a)(1+b) < 2$ équivaut à $a < \frac{1-b}{1+b}$; en choisissant a arbitrairement proche de $\frac{1-b}{1+b}$, on aura à la limite $2a + (2a+1)b = 2-b$ quantité qui est bien strictement plus grande que 1.

Pour avoir $2a \frac{(1+b)}{(1-b)} \geq 1$, il suffit donc de prendre a arbitrairement proche de $\frac{1-b}{1+b}$.

BIBLIOGRAPHIE

- [Bi 69] Billingsley "Convergence on probability measures" Wiley.
- [Br 68] Breiman "Probability" Addison-Wesley.
- [CeDi 84] Celeux, Diebolt "Reconnaissance de mélange de densités et classification. Un algorithme d'apprentissage probabiliste : l'algorithme SEM" Rapport de Recherche INRIA n° 349.
- [CeDi 85] Celeux, Diebolt "The SEM algorithm : a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem" Computational Statistics Quaterly, Vol 2, Issue 1.
- [CeDi 86] Celeux, Diebolt "L'algorithme SEM : un algorithme d'apprentissage probabiliste pour la reconnaissance de mélange de densités" Revue de Statistique Appliquée, Nov 86.
- [DLR 77] Dempster, Laird, Rubin "Maximum likelihood from incomplete data via the EM algorithm" JRSS.b. Vol 39.
- [ReWa 84] Redner, Walker "Mixture densities, maximum likelihood and the EM algorithm" SIAM Review, Vol 26, n° 2, April 84.
- [Si 80] Silvereman, "Some asymptotic properties of the probabilistic teacher" IEEE. Information theory, Vol. 26, n° 2.
- [Wu 83] Wu "On the convergence of the EM algorithm" Annals of Statistics 83, Vol, n° 1.

Imprimé en France

par

l'Institut National de Recherche en Informatique et en Automatique

