

STATISTIQUE ET ANALYSE DES DONNÉES

J. LEMAIRE

Propriétés asymptotiques en classification (consistance des solutions de problèmes approchés)

Statistique et analyse des données, tome 8, n° 1 (1983), p. 41-58

http://www.numdam.org/item?id=SAD_1983__8_1_41_0

© Association pour la statistique et ses utilisations, 1983, tous droits réservés.

L'accès aux archives de la revue « Statistique et analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

PROPRIETES ASYMPTOTIQUES EN CLASSIFICATION
(CONSISTANCE DES SOLUTIONS DE PROBLEMES APPROCHES)

J. LEMAIRE

Université de NICE
LASSY - ERA 835
41, bd Napoléon III - 06041 NICE CEDEX

Résumé : *La stabilité asymptotique des solutions de problèmes de classification se pose au praticien de l'analyse de données lorsque le nombre des observations à classer, n , est très grand. Du point de vue mathématique, ceci revient à étudier la convergence des solutions de ces problèmes lorsque n tend vers l'infini. Des résultats de convergence sont établis, sous des hypothèses très générale, dans le cas des méthodes centroïdes. L'outil de base est la notion d'intégrande normale introduite en analyse numérique.*

Abstract : *The asymptotic stability of the solutions of clustering problems arises in data analysis when the number, n , of the observations to be classified is very large. Mathematically speaking, this is equivalent with the study of convergence of the these solutions when n is being infinite. Convergence results are proved, with very general hypotheses, and concerning the centroïd methods. The main tool is the notion of normal integrand introduced in numerical analysis.*

Mots clés : *classification, méthodes centroïdes, propriétés asymptotiques, intégrande normale.*

1 - METHODES CENTROIDES. PROBLEME TRAITE

En classification automatique, on définit souvent les classes d'une partition à l'aide d'un ensemble de "centres" $y = \{m_1, \dots, m_l\}$.

C'est le cas, par exemple, lorsqu'on se propose de minimiser l'inertie intra-classe d'une partition $P = \{C_1, \dots, C_k\}$. Les observations étant situées dans un espace euclidien, ce critère s'écrit :

$$W(P) = \sum_j \sum_{i \in C_j} \|x_i - m_j\|^2$$

où m_1, \dots, m_k sont les centres de gravité des différentes classes et x_1, \dots, x_n les observations.

Lorsque $W(P)$ est minimum, il est bien connu que chaque observations x_i doit être affectée à la classe définie par le centre le plus proche. Plus précisément, les deux problèmes :

$$\inf_P W(P) \quad \text{et} \quad \inf_y \sum_i \inf_{m \in y} \|x_i - m\|^2$$

sont équivalents, ce qui montre bien qu'à l'optimum, les centres caractérisent les classes.

Cette remarque reste vraie pour tout critère additif du type :

$$\sum_j \sum_{i \in C_j} v(x_i, m_j)$$

car il est facile de constater que sa minimisation équivaut à celle du critère suivant :

$$\sum_i \inf_{m \in y} v(x_i, m)$$

Comme nouvel exemple, on peut citer le critère de la vraisemblance classifiante :

$$L(P) = \sum_j \sum_{i \in C_j} \text{Log } f(x_i, m_j)$$

où les $f(\cdot, m_j)$ sont des densités de probabilités. En posant :

$$v(x, m) = -\text{Log } f(x, m)$$

les deux problèmes :

$$\sup_P L(P) \text{ et } \inf_y \sum_i \inf_{m \in y} v(x_i, m)$$

sont équivalents.

Dans la suite, nous allons nous intéresser aux critères additifs du type précédent. Les centres m_1, \dots, m_k seront tous choisis dans un ensemble G dont nous préciserons ultérieurement la nature topologique. Une partition étant caractérisée par un ensemble d'au plus k centres, nous poserons :

$$F = \{y \subset G \mid \text{card } y \leq k\}$$

Dès à présent, observons que cette notion ensembliste permet de ne pas introduire un ordre non naturel entre les classes et élimine les problèmes de coïncidence de deux centres.

Pour simplifier les notations, nous poserons :

$$u(x, y) = \inf_{m \in y} v(x, m)$$

Disposant d'une suite d'observations $\underline{x} = (x_1, \dots, x_n)$, considérons le problème de la minimisation, par rapport à y , du critère suivant :

$$\hat{u}_n(\underline{x}, y) = (1/n) \sum_i u(x_i, y)$$

Dans cette étude, on se propose d'étudier la convergence des solutions de ces problèmes lorsque n tend vers l'infini. On montrera que cette convergence lieu vers l'ensemble des solutions du problème :

$$\inf_{y \in F} \hat{u}(y)$$

associé à un critère du type :

$$\hat{u}(x) = \int u(x, y) d\mu(x)$$

Cette convergence nécessite un cadre probabiliste qui sera précisé plus loin. Auparavant, mentionnons les résultats déjà acquis dans ce type d'étude asymptotique. POLLARD D. (1981) a considéré le cas du critère de l'inertie intra-classe. BRYANT P. et WILLIAMSON J.A. (1978) se sont intéressés à celui de la vraisemblance classifiante. Bien qu'utilisant des hypothèses spécifiques aux critères traités, ces études présentent des résultats analogues à ceux établis par WALD A. (1949) pour le critère du maximum de vraisemblance. Sur la base de ces analogies, nous nous sommes efforcés de mettre en place un cadre mathématique suffisamment général pour englober les études précédemment citées et certains de leurs prolongements, notamment ceux de REDNER R. (1981) et de PFANZAGL J. (1969), toujours sur le critère du maximum de vraisemblance. Dans souci de concision, nous nous limiterons à appliquer les résultats obtenus au cas des méthodes centroïdes en classification.

2 - HYPOTHESES TOPOLOGIQUES

Ces hypothèses vont permettre de préciser les notions de convergence dans l'espace des représentants de partitions. L'outil de base est la notion d'intégrande normale (H4) déjà largement utilisée en analyse numérique. Avec des hypothèses très faibles, cet outil assurera l'existence de solutions mesurables pour les problèmes d'optimisation posés.

E et G désignerons respectivement l'espace des observations et l'espace des centres. On rappelle que l'espace des représentants de partitions est défini par :

$$F = \{y \in G \mid \text{card } y \leq k\}$$

Dans toute la suite, on supposera :

H1 E et G sont des espaces topologiques métrisables, séparables et localement compacts

Il est facile de voir qu'il en est de même pour F lorsqu'on le munit de la topologie la plus fine rendant l'application suivante continue (topologie finale) :

$$\varphi: \underline{m} = (m_1, \dots, m_k) \in G^k \longrightarrow \{\underline{m}\} = \{m_1, \dots, m_k\}$$

PROPOSITION 1. Si d est une distance définissant la topologie de G, celle de F peut être définie par la distance de Hausdorff associée :

$$h(y, y') = \sup_{m \in y} \left(\inf_{m' \in y'} d(m, m') \right), \sup_{m' \in y'} \left(\inf_{m \in y} d(m, m') \right)$$

* La topologie définie par h rend l'application continue car, pour tout $\underline{m}, \underline{m}'$ dans G^k :

$$h(\{\underline{m}\}, \{\underline{m}'\}) \leq \sup_j d(m_j, m'_j)$$

Elle est donc contenue dans la topologie finale. Inversement, tout ouvert W' de la topologie finale, contenant y, contient une boule $B_h(y, r)$ car, pour r assez petit :

$$\varphi^{-1}(B_h(y, r)) \subset \bigcup_{\underline{m} \in \varphi^{-1}(y)} \prod_j B_d(m_j, r) \subset \varphi^{-1}(W')$$

Avant d'aborder les hypothèses suivantes, observons que l'hypothèse H1 nous permettra de manipuler aisément les notions de convergence, de mesurabilité et d'intégration sur E ou F. C'est d'ailleurs dans cette optique que nous l'avons formulée.

Considérons maintenant les hypothèses :

H2 μ mesure de Borel sur E

H3 (M_n) suite croissante de compacts recouvrant G

H4 ν intégrande normale sur $E \times G$

Par définition, X et Y étant deux espaces métrisables, séparables et localement compacts et ν une mesure de Borel sur X , une intégrande normale sur $X \times Y$ est une fonction w de $X \times Y$ dans $]-\infty, +\infty]$ pour laquelle il existe un borélien B de X , dont le complémentaire est ν -négligeable et tel que :

$\forall x \in B : w(x, \cdot)$ est semi-continue inférieurement sur Y
 w est borélienne sur $B \times Y$

Cette notion introduite en analyse numérique par BERLIOCCHI M. et LASSRY J.M. (1973) et développée notamment par EKELAND I. et TEMAM R. (1978) est particulièrement bien adaptée à notre problème. La semi-continuité inférieure de $v(x, \cdot)$ pour μ -presque tout x est une hypothèse naturelle dans les problèmes d'optimisation qui avait d'ailleurs été déjà proposée par PFANZAGL J. (1969) en visant des problèmes de maximisation de la vraisemblance.

L'hypothèse $H4$ est bien acquise lorsque $v(x, m) = ||x - m||^2$ et dans la plupart des cas pratiques lorsque $v(x, m) = -\text{Log } f(x, m)$. D'une manière plus générale, on peut montrer que toute fonction v de $E \times G$ dans $]-\infty, +\infty]$ telle que:

pour μ -presque tout $x : v(x, \cdot)$ est continue
pour tout $m : v(\cdot, m)$ est μ -mesurable

est une intégrande normale sur $E \times G$ (EKELAND I. et TEMAM R. (1978) par exemple).

Dans les démonstrations, la caractérisation suivante (dite de LUSIN) est particulièrement utile. X, Y et ν étant comme dans la définition qui précède, une fonction w de $X \times Y$ dans $]-\infty, +\infty]$ est une intégrande normale si et seulement si, pour tout compact I de X et tout $\varepsilon > 0$, il existe un compact $J \subset I$ tel que:

$\nu(I - J) \leq \varepsilon$
 w semi-continue inférieurement sur $J \times Y$

(référence précédente par exemple).

PROPOSITION 2. Avec les hypothèses H1, H2 et H4, la fonction u définie par:

$$u(x, y) = \inf_{m \in y} v(x, m)$$

est une intégrande normale sur $E \times F$.

* On utilise la caractérisation de LUSIN pour les intégrandes normales sur $E \times F$. Soient I un compact de E et $\varepsilon > 0$. Comme v est une intégrande normale, il existe un compact $J \subset I$ tel que :

$$\begin{aligned} \mu(I - J) &\leq \varepsilon \\ v &\text{ semi-continue inférieurement sur } J \times G \end{aligned}$$

Montrons qu'il en est de même pour u sur $J \times F$. Soient $(x, y) \in J \times F$ et (x_n, y_n) une suite dans $J \times F$ convergeant vers (x, y) . On commence par extraire une sous-suite (x_n', y_n') telle que :

$$\underline{\lim} u(x_n, y_n) = \lim u(x_n', y_n')$$

puis, pour tout n' , on choisit $m_n' \in y_n'$ tel que :

$$u(x_n', y_n') = v(x_n', m_n')$$

Comme la suite (y_n') converge vers y , on peut à nouveau extraire un sous-suite (m_n'') convergeant vers $m \in y$. La fonction v étant semi-continue inférieurement sur $J \times G$, on peut écrire :

$$\underline{\lim} v(x_n'', m_n'') \geq v(x, m)$$

Mais, comme:

$$v(x, m) \geq u(x, y)$$

on obtient bien l'inégalité recherchée:

$$\underline{\lim} u(x_n, y_n) \geq u(x, y)$$

La propriété suivante des intégrandes normales (existence d'une fonction minimisante) sera utilisée. Avec les notations de la définition, pour tout compact K de Y , il existe une fonction ν -mesurable s de X dans K telle que, pour ν -presque tout x :

$$w(x, s(x)) = \inf_{y \in K} w(x, y)$$

(EKELAND I. et TEMAM R. (1978) par exemple). Afin de bien fixer les notions utilisées, précisons, pour clore ce paragraphe que la ν -mesurabilité de s signifie que $s^{-1}(B(K)) \subset B_\nu(X)$ où $B(X)$ désigne la tribu borélienne de X et $B_\nu(X)$, la tribu complétée de $B(X)$ relativement à la mesure ν .

PROPOSITION 3. Avec les hypothèses H1 à H4, pour tout entier n , il existe une fonction μ^n -mesurable s_n de E^n dans le compact $K_n = \{y \in F \mid y \subset M_n\}$ telle pour μ^n -presque tout $x \in E^n$:

$$\hat{u}_n(\underline{x}, s_n(\underline{x})) = \inf_{y \subset M_n} \hat{u}(\underline{x}, y)$$

* \hat{u}_n est une intégrande normale sur $E^n \times F$. Cela résulte de la proposition 2 et du fait que la tribu borélienne de E^n est le produit de n tribus boréliennes de E puisque la topologie de E admet une base dénombrable. Il suffit alors d'utiliser le résultat précédemment énoncé sur la minimisation des intégrandes normales. *

3 - HYPOTHESES D'INTEGRABILITE

Les hypothèses qui suivent visent, avec des conditions très faibles d'intégrabilité, à permettre l'application ultérieure de la loi des grands nombres pour établir les résultats de convergence. Ces hypothèses visent aussi à localiser dans un compact les solutions du problème limite (proposition 7).

Grâce aux hypothèses et aux résultats du paragraphe précédent, pour tout compact M de G et tout compact K de F , les fonctions:

$$\inf_{m \in M} v(., m) \text{ et } \inf_{y \in K} u(., y)$$

sont μ -mesurables. Il en sera donc de même si M ou K sont des réunions dénombrables de compacts, par exemple des ouverts.

H5 Pour tout $m \in G$, il existe un voisinage compact T de m tel que:

$$\int_{m' \in T} (\inf v(x, m'))^- d\mu(x) < +\infty \text{ (quasi-intégrabilité).}$$

PROPOSITION 4. Avec les hypothèses H1 à H5, pour tout $y \in F$, il existe un voisinage compact W de y tel que $\int_{y' \in W} (\inf u(x, y'))^- d\mu(x) < +\infty$.

* Si $y \in F$, il existe $r > 0$ tel que, pour tout $m \in y$, $B_d(m, r)$ est compacte et :

$$\int_{m' \in B_d(m, r)} (\inf v(x, m'))^- d\mu(x) < +\infty$$

Dès lors, $W = B_h(y, r)$ est un voisinage compact de y tel que :

$$\int_{y' \in W} (\inf u(x, y'))^- d\mu(x) \leq \sum_{m \in y} \int_{m' \in B_d(m, r)} (\inf v(x, m'))^- d\mu(x)$$

En particulier, pour tout $y \in F$, on peut donc bien définir :

$$\hat{u}(y) = \int u(x, y) d\mu(x)$$

puisque la fonction $u(\cdot, y)$ est μ -mesurable (proposition 2) et que :

$$\int u(x, y) d\mu(x) \leq \int_{y' \in W} (\inf u(x, y'))^- d\mu(x) < +\infty$$

PROPOSITION 5. Avec les hypothèses H1 à H5, la fonction \hat{u} est semi-continue inférieurement sur F .

* Soient $y \in F$, (y_n) une suite dans F convergeant vers y et W , un voisinage compact de y satisfaisant à l'inégalité de la proposition précédente. Pour n assez grand, $y_n \in W$ et donc la fonction $u(\cdot, y_n)$ est minorée par la fonction quasi intégrable $(\inf u(x, y'))^-$. D'après le lemme de Fatou Lebesgue, on obtient :

$$\liminf \int u(x, y_n) d\mu(x) \geq \int \liminf u(x, y_n) d\mu(x)$$

Par ailleurs, comme $u(x, \cdot)$ est semi-continue inférieurement pour μ -presque tout x , on peut affirmer que :

$$\int \liminf u(x, y_n) d\mu(x) \geq \int u(x, y) d\mu(x)$$

On vient de prouver l'inégalité visée: $\liminf \hat{u}(y_n) \geq \hat{u}(y)$.

Posons alors $\hat{u} = \inf_{y \in F} \hat{u}(y)$ et formulons les hypothèses :

$$H6 \quad \hat{u} < +\infty$$

et, lorsque G n'est pas compact :

$$H7 \quad a) \exists M \text{ compact } \subset G : \int (\inf_{m \in G-M} v(x, m))^- d\mu(x) < +\infty$$

$$b) \text{ pour } \mu\text{-presque tout } x \in E : \lim_{m \rightarrow \infty} v(x, m) = +\infty$$

La proposition suivante est le résultat clef dans la démonstration de la convergence annoncée au paragraphe 1.

PROPOSITION 6. Avec les hypothèses H1 à H6 et H7 si G n'est pas compact, pour tout $y \in F$ vérifiant $\hat{u} < \hat{u}(y)$, il existe un voisinage compact W de y et un compact $M \subset G$ tel que la fonction $\inf\{\inf_{z \in W} u(\cdot, z), \inf_{m \in G-M} v(\cdot, m)\}$ soit quasi-intégrable et vérifie :

$$\hat{u} < \int \inf_{z \in W} \{ \inf_{m \in G-M} u(x, z), \inf_{m \in G-M} v(x, m) \} d\mu(x)$$

Il existe donc un compact $M \subset G$ tel que la fonction $\inf_{m \in G-M} v(\cdot, m)$ soit quasi-intégrable et vérifie :

$$\hat{u} < \int \inf_{m \in G-M} v(x, m) d\mu(x)$$

* On raisonne dans le cas non compact en laissant au lecteur le soin de supprimer tout ce qui est inutile dans le cas compact. Soient (W_n) un système fondamental de voisinages compacts de y et (M_n) une suite exhaustive de

compacts de G. Pour n assez grand, les fonctions μ -mesurables :

$$f_n = \inf_{z \in W_n} (\inf u(.,z), \inf_{m \in G-M_n} v(.,m))$$

sont donc minorées par une fonction quasi-intégrable (hypothèse H7a et proposition 6). Elle sont donc quasi-intégrables. Par ailleurs, la suite (f_n) est une suite croissante. d'après le théorème de convergence monotone de Lebesgue, on a donc :

$$\lim \int f_n(x) d\mu(x) = \int \lim f_n(x) d\mu(x)$$

Pour conclure, il suffit de remarquer que, pour μ -presque tout x, $f_n(x)$ converge vers $u(x,y)$. Cela résulte de la semi-continuité inférieure de $u(x,.)$ et de l'hypothèse H7b. Comme $\bar{u} < \hat{u}(y)$, il suffit alors de prendre $W = W_n$ et $M = M_n$ pour n assez grand. *

Pour clore ce paragraphe, il reste à préciser la structure de l'ensemble $H = \{y \in F \mid \bar{u}(y) = \hat{u}.\}$, l'ensemble des solutions du problème $\min \hat{u}(y)$. Lorsque G n'est pas compact, on supposera :

$$H8 \quad \bar{u} < \inf \{ \hat{u}(y) \mid y \in F \text{ et } \text{card } y < k \}$$

On notera que, d'après la forme de u, $k' \leq k'' \leq k$ implique toujours :

$$\inf \{ \hat{u}(y) \mid y \in F \text{ et } \text{card } y \leq k' \} \leq \inf \{ \hat{u}(y) \mid y \in F \text{ et } \text{card } y \leq k'' \}$$

L'hypothèse H8 est moins restrictive que celle utilisée par POLLARD D. (1981) qui suppose, pour chaque $k' \leq k$, l'existence d'une solution unique aux problèmes du type précédent et la stricte décroissance des minima, ceci afin de permettre une démonstration par récurrence.

PROPOSITION 7. Avec les hypothèses H1 à H6 et H7, H8 si G n'est pas compact, H est un ensemble compact non vide.

* Comme la fonction \bar{u} est semi-continue inférieurement (proposition 5), il suffit de montrer que :

$$\exists K \text{ compact } \subset F : \hat{u} . < \inf_{y \in F - K} \hat{u}(y)$$

C'est évident si G (donc F) est compact car on peut prendre $K=F$. Dans le cas contraire, on raisonne par l'absurde. Si ce n'est pas vrai, il existe une suite (y_n) dans F telle que :

$$\lim y_n = \infty \text{ et } \hat{u} . = \lim \hat{u}(y_n)$$

D'abord, on peut exclure le cas où toute suite (m_n) prélevée dans (y_n) converge vers ∞ (on dira alors que la suite (y_n) converge totalement vers ∞). En effet, pour tout compact M de G et tout entier n assez grand, on aurait :

$$\hat{u}(y_n) \geq \int_{m \in G - M} \inf v(x, m) d\mu(x)$$

Comme on peut trouver de tels compacts M tels que :

$$\hat{u} . < \int_{m \in G - M} \inf v(x, m) d\mu(x)$$

(proposition 6), on contredirait la convergence de la suite $(\hat{u}(y_n))$ vers $\hat{u} .$

On peut donc prélever dans la suite (y_n) à la fois des suites (m_n) convergeant vers ∞ et des suites (m'_n) ne convergeant pas vers ∞ c'est à dire contenant une sous-suite à valeurs dans un compact de G. Par extractions successives de sous-suites, on peut alors conclure à l'existence d'une sous-suite (y_n') et, pour tout n d'une partition de y_n' en deux classes non vides z_n' et t_n' telles que :

la suite (z_n') converge totalement vers ∞ ,

la suite (t_n') converge vers $y \in F$ tel que $\text{card } y < k$.

Comme $\hat{u} . < \hat{u}(y)$ (hypothèse H8), la proposition 7 permet d'affirmer l'existence d'un voisinage compact W de y et d'un compact M de G tels que :

$$\hat{u} . < \int_{z \in W} \inf (\inf u(x, z), \inf_{m \in G - M} v(x, m)) d\mu(x)$$

la fonction intégrée étant quasi-intégrable. Pour n assez grand, $z_n \in G-M$ et $t_n \in W$. On peut donc affirmer que, pour de tels n :

$$\int \inf_{z \in W} (\inf u(x,z), \inf_{m \in G-M} v(x,m)) d\mu(x) \leq \hat{u}(y_n')$$

ce qui, là encore contredit la convergence de la suite $(\hat{u}(y_n))$ vers \hat{u} .

4 - HYPOTHESES PROBABILISTES ET RESULTATS DE CONVERGENCE

On revient maintenant aux problèmes approchés :

$$\inf_{y \in M_n} \hat{u}_n(\underline{x}, y) \quad \text{où} \quad \hat{u}_n(\underline{x}, y) = (1/n) \sum_i u(x_i, y)$$

Pour que $\hat{u}_n(\underline{x}, y)$ puisse être considéré comme une approximation de $\hat{u}(y)$, il suffit, grâce à la loi des grands nombres, de supposer que \underline{x} est la réalisation tronquée d'une suite (X_n) de variables aléatoires, indépendantes et de même loi μ . C'est l'objet de la dernière hypothèse formulée qui peut être considérée comme un passage à la limite des cas où un grand nombre d'observations indépendantes doivent être classifiées.

H9 (X_n) est une suite de variables aléatoires, définies sur un espace probabilisé complet (Ω, \mathcal{A}, P) , à valeurs dans E , indépendantes et de même loi μ .

Là encore, pour bien fixer les notions manipulées, il convient de préciser qu'une variable aléatoire X de (Ω, \mathcal{A}, P) dans un espace topologique θ est une application de Ω dans θ telle que $X^{-1}(B(\theta)) \subset \mathcal{A}$, $B(\theta)$ désignant la tribu borélienne de θ . Comme la tribu \mathcal{A} est P -complète, on peut remplacer dans cette inclusion $B(\theta)$ par $B_\lambda(\theta)$, la tribu complétée relativement à λ , la loi de X .

THEOREME 1. Avec les hypothèses H1 à H4 et H9, pour tout entier n , il existe une variable aléatoire Y_n de (Ω, \mathcal{A}, P) dans F telle que, pour presque tout $\omega \in \Omega$:

$$\hat{u}_n(\underline{x}_n(w), Y_n(w)) = \inf_{y \in M_n} \hat{u}_n(\underline{x}_n(w), y)$$

où $\underline{x}_n = (x_1, \dots, x_n)$. Une telle suite (Y_n) est dite minimisante.

* Avec les notations de la proposition 3, il suffit de prendre :

$$Y_n(w) = s_n(\underline{x}_n(w))$$

On obtient bien une variable aléatoire car la fonction s_n est μ^n -mesurable et \underline{x}_n est une variable aléatoire de loi μ^n .

THEOREME 2. Avec les hypothèses H1 à H6, H9 et H7, H8 si G n'est pas compact, toute suite (Y_n) minimisante converge presque sûrement vers H, l'ensemble des solutions du problème $\min_{y \in F} \hat{u}(y)$.

$$y \in F$$

* F étant un espace localement compact, métrisable et séparable, il existe un ensemble dénombrable Φ , constitué de couples (W, M) où W est un compact de F et M un compact de G tels que la fonction $\inf_{z \in W} (\inf_{m \in M} u(\cdot, z), \inf_{m \in M} v(\cdot, m))$ soit quasi-intégrable, et dans lequel on peut choisir les ensembles W et M de la proposition 6 pour tout $y \in F$ tel que $\hat{u} < \hat{u}(y)$. Par ailleurs, soit M_0 un compact de G tel que la fonction $\inf_{m \in M_0} v(\cdot, m)$ soit quasi-intégrable et vérifie:

$$\hat{u} < \int_{m \in G - M_0} \inf v(x, m) d\mu(x)$$

Considérons enfin un élément y de H. D'après la loi des grands nombres sur les variables aléatoires quasi-intégrables, pour P-presque tout $w \in \Omega$, on peut affirmer que :

$$\lim \hat{u}_n(\underline{x}_n(w), y) = \hat{u}.$$

$$\lim (1/n) \sum_{i=1, n} \inf_{m \in G - M_0} v(x_i(w), m) = \int_{m \in G - M_0} \inf v(x, m) d\mu(x)$$

$$\begin{aligned} \forall (W, M) \in \Phi : \lim (1/n) \sum_{i=1, n} \inf_{z \in W} (\inf_{m \in M} u(x_i(w), z), \inf_{m \in M} v(x_i(w), m)) \\ = \int_{z \in W} \inf_{m \in M} (\inf u(x, z), \inf v(x, m)) d\mu(x) \end{aligned}$$

Choisissons $w \in \Phi$ réalisant ces convergences et tel que, pour tout n :

$$\hat{u}_n(X_n(w), Y_n(w)) = \inf_{y \in M_n} \hat{u}_n(X_n(w), y)$$

(presque tout w convient donc.)

Soit $y_n = Y_n(w)$. Prouvons d'abord l'existence d'un compact K de F contenant tous les termes de la suite (y_n) . C'est évident si G (donc F) est compact. Dans le cas contraire, on raisonne par l'absurde comme dans la démonstration de la proposition 7. D'abord, on élimine le cas où la suite (y_n) converge totalement vers ∞ car, pour n assez grand, y_n serait contenu dans $G - M_0$ ce qui, grâce aux convergences précédentes, impliquerait $\hat{u}_n(X_n(w), y.) < \hat{u}_n(X_n(w), y_n)$. Or cette inégalité est impossible dès que $y. \in M_n$. On peut donc extraire une sous-suite $(y_{n'})$ dont chaque terme $y_{n'}$ peut être partitionné en deux classes non vides $z_{n'}$ et $t_{n'}$ telles que :

- la suite $(z_{n'})$ converge totalement vers ∞ ,
- la suite $(t_{n'})$ converge vers $y \in F$ avec $\text{card } y < k$.

Comme $\hat{u}. < \hat{u}(y)$ d'après l'hypothèse H8, il existe un couple $(w, M) \in \Phi$ vérifiant l'inégalité de la proposition 6. Pour n assez grand, dès que $z_{n'} \in W$ et $t_{n'} \in M$, en utilisant les convergences précédentes, on peut donc affirmer que :

$$\hat{u}_{n'}(\underline{X}_{n'}(w), y.) < \hat{u}_{n'}(\underline{X}_{n'}(w), y_{n'})$$

ce qui est impossible lorsque $y. \in M_{n'}$. On aboutit donc à une contradiction dans tous les cas.

Montrons maintenant la convergence annoncée. Elle signifie que, pour tout voisinage V de H , il existe un entier n_0 tel que, pour tout $n \geq n_0$, on ait $y_n \in V$. Soit donc V un tel voisinage supposé ouvert. Reprenons l'argumentation de WALD A. (1949). Pour tout $y \in K - V$, on a $\hat{u}. < \hat{u}(y)$. Il existe donc un couple $(w, M) \in \Phi$ vérifiant l'inégalité de la proposition 6 où W est un voisinage compact de y . Comme $K - V$ est compact, il existe un recouvrement fini de $K - V$ par de tels voisinages W_1, \dots, W_q . Pour tout n on a donc :

$$\begin{aligned}
 & \hat{u}_n(\underline{X}_n(w), y) - \inf_{y \in K-V} \hat{u}_n(\underline{X}_n(w), y) \\
 & \leq \hat{u}_n(\underline{X}_n(w), y) - \inf_j \inf_{y \in W_j} \hat{u}_n(\underline{X}_n(w), y) \\
 & \leq \hat{u}_n(\underline{X}_n(w), y) - \inf_j (1/n) \sum_{i=1, n} \inf_{y \in W_j} (\inf_{m \in G-M_j} u(X_i(w), y), \inf_{m \in G-M_j} v(X_i(w), m)) \\
 & \leq \sup_j (1/n) \sum_{i=1, n} (u(X_i(w), y) - \inf_{y \in W_j} (\inf_{m \in G-M_j} u(X_i(w), y), \inf_{m \in G-M_j} v(X_i(w), m))) \\
 & < 0
 \end{aligned}$$

pour tout n assez grand d'après les convergences précédentes. Comme tous les termes de la suite (y_n) sont dans K, il suffit de choisir n_0 tel que $y \subset M_n$ *

Remarque : cette convergence équivaut à dire que :

$$\lim_{y \in H} \inf h(Y_n(w), y) = 0 \text{ presque sûrement}$$

5 - CONCLUSION

Le cadre topologique et probabiliste que nous venons de définir permet d'étudier la stabilité asymptotique des solutions de certains problèmes de classification. Cette stabilité est acquise sous des hypothèses très générales dans le cas des méthodes centroïdes lorsqu'on recherche des optima globaux.

Plusieurs problèmes naturels se posent alors.

Les optima globaux étant des notions difficilement manipulables en classification, peut-on étendre ces résultats à des méthodes centroïdes non globalement optimales ?

Peut-on apprécier la rapidité de la convergence démontrée ?

Cette étude peut-elle s'appliquer à d'autres types de méthodes de classification ?

Nous possédons déjà des réponses partielles à ces questions. Par exemple, le cadre ensembliste a pu être étendu à des méthodes centroïdes proches de celles étudiées notamment par MAC QUEEN J.(1967). Des résultats de convergence analogues à ceux du paragraphe précédent ont pu être établis en adaptant à ce cadre ensembliste des techniques d'approximation stochastique (PHILIP H. (1983)). Parallèlement, nous avons pu étendre ce type de résultats au problème de l'estimation des composants d'un mélange de lois de probabilités.

Il semble donc bien que des études de propriétés asymptotiques en classification puissent être développées dans plusieurs directions.

6 - BIBLIOGRAPHIE

BERLIOCHI H. et LASSRY J.M. (1973)

Intégrales normales et mesures paramétrées en calcul des variations. Bull. soc. math. France. 101, p 129-184.

BRYANT P. et WILLIAMSON J. (1978)

Asymptotic behaviour of classification maximum likelihood estimates. Biometrika. 65, 2, p 273-281.

EKELAND I. et TEMAM R. (1978)

Analyse convexe et problèmes variationnels. Dunod. Paris

MAC QUEEN J. (1967)

Some methods for classification and analysis of multivariate observations. Proc. 5th Berkeley symp. 1, p 281-297.

PHILIP H. (1983)

Propriétés asymptotiques en classification: convergence d'un schéma d'approximation stochastique. Thèse de 3e cycle. Université de Nice. (à paraître en janvier).

PFANZAGL J. (1969)

On the measurability and consistency of minimum contrast estimates. Metrika. 1., p 249-272.

POLLARD D.(1981)

Strong consistency of k-means clustering. Ann. Stat. 9, 1, p 135-140.

REDNER R.(1981)

Note on the consistency of the maximum likelihood estimate for non identifiable distributions. 9, 1, p 225-228.

WALD A.(1949)

Note on the consistency of the maximum likelihood estimate. Ann. Math. Stat. 20, p 595-601

