

STATISTIQUE ET ANALYSE DES DONNÉES

JEAN-PIERRE ASSELIN DE BEAUVILLE

L'estimation des modes d'une densité de probabilité multidimensionnelle

Statistique et analyse des données, tome 8, n° 3 (1983), p. 16-40

http://www.numdam.org/item?id=SAD_1983__8_3_16_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/>

L'ESTIMATION DES MODES D'UNE DENSITE
DE PROBABILITE MULTIDIMENSIONNELLE

Jean-Pierre ASSELIN de BEAUVILLE

Laboratoire d'Informatique
U.E.R. A.G.I.
Parc Grandmont - 37200 TOURS

Résumé : Cet article décrit une méthode d'estimation des régions modales ou des modes d'une loi de probabilité admettant une densité multidimensionnelle. C'est une généralisation d'un travail précédent relatif au cas unimodal et univarié [1]. La méthode utilisée est très simple : Sur un histogramme traditionnel on repère les régions de densité maximum à l'aide d'un seuil.

On donne les conditions de convergence presque sûre de l'algorithme. On notera que, par rapport à l'algorithme de SAGER [18], on s'affranchit ici de la contrainte de convexité des régions modales.

Abstract : This paper presents a method for estimating modal sets or modes of a multimodal and multivariate density. This work generalizes a previous paper on univariate mode's estimation [1]. We use a simple idea : With a classical histogram-type estimator, we select the higher density sets with the help of a moving density threshold.

The estimator is shown to converge almost surely. Compared with SAGER's algorithm [18] it is interesting to remark that here, the modal sets need not to be convex.

Mots clés : Estimation du mode, densité multivariée, recherche des groupements.

1 - INTRODUCTION

L'objet de cette note est d'étendre aux distributions multidimensionnelles et multimodales l'algorithme d'estimation du mode d'une loi unidimensionnelle et unimodale proposé par l'auteur dans [1].

Suivant le cas, ce problème peut se présenter sous deux aspects :

Soit celui de l'estimation des régions modales dans le but, par exemple, d'effectuer une classification automatique,

Soit celui de l'estimation des modes de la loi considérée.

Dans la seconde éventualité, on ne pourra pas en général fournir une estimation simultanée des modes. Par contre, une fois obtenu un nombre stable de régions unimodales, on pourra, à l'aide du même estimateur, estimer successivement les modes de ces différentes régions.

L'estimation du mode d'une distribution univariée a fait l'objet de nombreuses publications parmi lesquelles on peut citer : [2] à [14].

L'estimation du (ou des) mode(s) d'une loi multivariée a été étudiée essentiellement par SAMANTA [15], KONAKOV [16] et SAGER [17] et [18].

SAMANTA [15] adopte la méthode proposée par PARZEN [2] en la généralisant au cas multivarié : l'estimateur $\hat{\theta}_n$ (θ_n est un vecteur de R^p) est le mode d'un estimateur à noyau de CACOULOS [19] de la densité de probabilité. Sous certaines conditions imposées à la densité, SAMANTA prouve la convergence presque sûre de $\hat{\theta}_n$ ainsi que sa normalité asymptotique.

SAGER [17] généralise le procédé utilisé par VENTER [7] et CHERNOFF [3] pour le cas univarié : à partir d'une classe de formes géométriques simples (des hypercubes ou hypersphères par exemple), SAGER propose d'estimer le mode $\hat{\theta}_n$ et donne un ordre de grandeur de sa vitesse de convergence.

Dans l'article de 1979, SAGER [18] propose d'estimer le mode en choisissant un point de l'ensemble obtenu à l'issue d'un processus itératif générant une suite décroissante d'ensembles convexes imbriqués les uns dans les autres. Il démontre la convergence presque sûre de l'estimateur et donne un ordre de grandeur de sa vitesse de convergence dans le cas où la loi est unidimensionnelle.

L'estimation des régions modales dans l'espace à p dimensions a été plus souvent développée en relation avec la classification automatique. En effet, on peut interpréter ces régions comme des estimations des sous-populations à distinguer dans la population soumise à l'étude. Sans prétendre

à l'exhaustivité on peut citer comme exemple de ce type d'approche : [20] à [33].

Dans la plupart des cas on réalise une estimation des régions unimodales qui sont ensuite utilisées comme noyaux pour former les taxons (clusters). Un des avantages de cette façon de procéder est qu'aucune réduction de variance n'est imposée aux classifications qui peuvent alors présenter différentes formes (sphérique, allongée, irrégulière, ramifiée, etc...).

L'estimateur décrit dans cet article se caractérise par sa simplicité. Comme SAMANTA [15], on utilise un estimateur de la densité mais dans sa forme la plus simple puisqu'il s'agit de l'histogramme classique. D'autre part, comme SAGER [18], on procède de façon itérative par l'obtention de suites décroissantes d'ensembles inbriqués mais sans que la contrainte de convexité de ces ensembles ne soit nécessaire à la démonstration de la convergence presque sûre. Suivant le but recherché, l'algorithme proposé pourra servir soit à l'estimation des modes de la densité, soit à l'estimation des régions modales dans le but d'effectuer ensuite, par exemple, une classification automatique.

2 - L'ALGORITHME

Soit $F(x_1, x_2, \dots, x_p) = F(\underline{x})$ une fonction de répartition absolument continue (par rapport à la mesure de Lebesgue) définie sur R^p et admettant $p+1$ dérivées continues et bornées. On notera $f(\underline{x})$ la fonction de densité correspondante :

$$f(\underline{x}) = \partial^p F(\underline{x}) / \partial x_1 \partial x_2 \dots \partial x_p .$$

On supposera aussi $f(\underline{x})$ continue et qu'elle possède $t (\geq 1)$ modes \underline{x}_j^0 , $j = 1, 2, \dots, t$ ($\underline{x}_j^0 = (x_{1j}^0, x_{2j}^0, \dots, x_{nj}^0)$) (la valeur exacte de t est supposée inconnue). En outre, on admettra qu'autour de chacun des modes, $f(\cdot)$ est strictement unimodale, c'est-à-dire que chaque mode \underline{x}_j^0 est un singleton et que dans le voisinage de \underline{x}_j^0 , $f(\cdot)$ est strictement décroissante le long de toute demi-droite de R^p issue de \underline{x}_j^0 . Il est clair

que pour $t > 1$, la stricte unimodalité de $f(\cdot)$ n'est vraie que localement autour des modes et pas globalement. De façon plus précise, on dira que \underline{x}^0 , appartenant à un sous-ensemble A' de \mathbb{R}^p , est un mode de $f(\cdot)$ si les deux propriétés suivantes sont vérifiées :

$$+ f(\underline{x}^0) > 0$$

+ Quel que soit $\underline{y} \in A'$, tel que $\underline{y} = \underline{x}^0 + \underline{\varepsilon}$ où

$$\underline{\varepsilon} = (\gamma_1 \varepsilon_1, \gamma_2 \varepsilon_2, \dots, \gamma_p \varepsilon_p), \quad \varepsilon_i \in \mathbb{R}, \quad i = 1, 2, \dots, p$$

et les γ_i prenant indépendamment les uns des autres les valeurs 0 ou 1 avec la contrainte $\sum_{i=1}^p \gamma_i > 0$, alors :

$$f(\underline{y}) < f(\underline{x}^0).$$

D'autre part, on supposera qu'autour de chaque mode \underline{x}_j^0 il existe un sous-ensemble non vide de \mathbb{R}^p désigné par A_j tel que quel que soit \underline{x} de A_j et quel que soit $\lambda \in]0, 1[$ on ait :

$$\underline{x}_1 = \lambda \underline{x}_j^0 + (1-\lambda) \underline{x} \in A_j \quad (A_j \text{ est étoilé en } \underline{x}_j^0)$$

et

$$f(\underline{x}) < f(\underline{x}_1) < f(\underline{x}_j^0) \tag{1}$$

Ainsi à chaque mode \underline{x}_j^0 de $f(\cdot)$ on suppose attaché un ensemble A_j constitué par le domaine de stricte décroissance de la densité autour de \underline{x}_j^0 .

Il faut noter que l'existence des A_j n'implique pas contrairement à l'une des hypothèses de SAGER, que la fonction $f(\cdot)$ ne présente pas de plateaux (régions de \mathbb{R}^p où $f(\cdot)$ reste constante). On se place donc ici dans un cas plus général. (Tout ce que l'on suppose ici c'est l'existence d'un domaine A_j autour de chaque mode \underline{x}_j^0 dans lequel $f(\cdot)$ ne présente pas de plateaux, mais ces plateaux peuvent évidemment exister en dehors des A_j).

On suppose enfin que la densité $f(\cdot)$ est telle qu'il existe au moins un réel $U > 0$ tel que l'ensemble $\{\underline{x} : f(\underline{x}) \geq U\}$ soit un ensemble borné, formé de la réunion d'un nombre fini d'ensembles connexes. Pour la suite on choisira un nombre $\sigma > 0$ appartenant à l'ensemble de u tels que

$\{\underline{x} : f(\underline{x}) \geq u\}$ est borné. σ sera choisi assez petit pour que tous les modes \underline{x}_j^0 de f soient dans l'ensemble borné $\{\underline{x} : f(\underline{x}) \geq \sigma\}$. On déduit de cette dernière hypothèse que les A_j sont nécessairement bornés.

Soient $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ ($\underline{x}_i = (x_{1i}, x_{2i}, \dots, x_{pi})$) un échantillon aléatoire de taille n issu de $F(\cdot)$. Soit $F_n(\cdot)$ la fonction de répartition empirique associée à cet échantillon :

$$n F_n(\underline{x}) = \text{nombre d'observations } \underline{x}_i, i = 1, 2, \dots, n \text{ telles que } x_{ji} < x_j \text{ pour tout } j = 1, 2, \dots, p.$$

Soit $(S(n))$ (ou S_n) une suite positive strictement décroissante. Pour chaque valeur de $n \in \mathbb{N}$ on définit une partition de l'espace \mathbb{R}^p en une suite infinie d'hypercubes de volume S_n^p . Pour cela chacune des p composantes de l'espace \mathbb{R}^p est partagée en une suite d'intervalles égaux de longueur S_n suivant le modèle :

$$\dots, [-2S_n, -S_n[, [-S_n, 0[, [0, S_n[, [S_n, 2S_n[, \dots$$

Soit $h(n, \underline{x})$ l'histogramme construit à partir de cette partition de \mathbb{R}^p :

$$h(n, \underline{x}) = (\Delta^{(1)} \dots \Delta^{(p)}) F_n[(q_1-1)S_n, \dots, (q_p-1)S_n] / S_n^p$$

où \underline{x} est un point de l'hypercube :

$${}^p q_1 \dots q_p = \begin{cases} (q_1-1)S_n \leq x_1 < q_1 S_n \\ \vdots \\ (q_p-1)S_n \leq x_p < q_p S_n \end{cases}$$

et $q_j \in \mathbb{Z}, j = 1, 2, \dots, p,$

$$\begin{aligned} \Delta^{(j)} F_n[(q_1-1)S_n, \dots, (q_p-1)S_n] &= F_n[(q_1-1)S_n, \dots, q_j S_n, \dots, (q_p-1)S_n] - \dots \\ &\dots - F_n[(q_1-1)S_n, \dots, (q_p-1)S_n] \end{aligned}$$

et où le produit des $\Delta^{(j)}$ signifie qu'il faut effectuer les opérations les unes après les autres en tenant compte de :

$$\Delta(F_n(\dots) - G_n(\dots)) = \Delta F_n(\dots) - \Delta G_n(\dots) .$$

Dans la suite on posera :

$$(q-1)S = ((q_1-1)S, \dots, (q_p-1)S) .$$

Considérons la suite aléatoire $(r(m,n))$ définie par le processus suivant :

- 1) Définir un nombre réel $l > 1$ (voir plus bas pour le choix de l).
- 2) Poser $m = 1$ et $L = 1/l$.
- 3) Calculer $r(0,n) = L \sup_{\underline{x} \in \mathbb{R}^p} h(n, \underline{x})$.
- 4) Si $\sup_{\underline{x} \in \mathbb{R}^p} h(n, \underline{x}) - r(m-1, n) > L$: aller en 5 ,
 Si non : $L = L/2$ et aller en 4 .
- 5) Calculer $r(m,n) = r(m-1, n) + L$.
- 6) Poser $m = m+1$ et aller en 4 .

L'algorithme d'estimation des régions modales de $f(\cdot)$ procède alors comme suit : Partant de $m = 1$ on associe à chaque valeur de $r(m,n)$ au moins un ensemble $C(r(m,n), w)$ tel que les deux propositions ci-dessous soient simultanément vérifiées :

$$\text{Pour tout } \underline{x} \text{ appartenant à un des } C(r(m,n), w) : h(n, \underline{x}) \geq r(m,n) \quad (2)$$

$$\text{Pour tout } \underline{x} \text{ n'appartenant à aucun des } C(r(m,n), w) : h(n, \underline{x}) < r(m,n) \quad (3)$$

On désignera par t_{mn} le nombre d'ensembles $C(r(m,n), w)$ obtenus pour le seuil $r(m,n)$, ce qui implique que l'indice w (attaché à chaque composante connexe de l'ensemble : $\{\underline{x} : h(n, \underline{x}) \geq r(m,n)\}$) prendra ses valeurs dans $\{1, 2, \dots, t_{mn}\}$ avec $1 \leq t_{mn}$. En faisant croître m , on trouvera donc, pour chaque valeur de $r(m,n)$, un certain nombre d'ensembles $C(\cdot, \cdot)$ à l'intérieur desquels $h(\cdot, \cdot)$ sera supérieur ou égal à $r(m,n)$. La mesure de ces ensembles diminuant au fur et à mesure que m croît. Ces ensembles $C(\cdot, \cdot)$ sont les estimations cherchées des régions modales.

Pour estimer le(s) mode(s) de $f(\cdot)$, il suffira d'appliquer le même algorithme successivement aux observations de chacun des ensembles $C(\cdot, \cdot)$. En faisant croître m , on pourra convenir d'estimer un mode, chaque fois

que l'effectif du $C(r(m,n),w)$ considéré devient inférieur à un seuil n_s ($0 < n_s < n$) fixé à l'avance, par un point \underline{x}_{wn}^0 intérieur à l'ensemble $C(r(m,n),w)$ étudié (si cet ensemble contient plus d'un mode, c'est le plus dense qui sera estimé).

Les frontières des ensembles $C(.,.)$ sont, pour m et n donnés, des estimations des courbes d'iso-densité de $f(.)$.

3 - CONVERGENCE ASYMPTOTIQUE

Considérons la suite $(r(m))$ définie de la façon suivante :

- 1) Définir un nombre réel l supérieur à un (on donne à l la même valeur que pour $r(m,n)$).
 - 2) Poser $m = 1$ et $L = 1/l$
Poser $r(0) = L \cdot \sup_{\underline{x} \in R^p} f(\underline{x})$.
 - 3) Si $\sup_{\underline{x} \in R^p} f(\underline{x}) - r(m-1) > L$: Aller en 4 ,
Si non : $L/2$ et aller en 3 .
 - 4) $r(m) = r(m-1) + L$.
 - 5) $m = m+1$ et aller en 3 .
- On peut énoncer le lemme 1 :

Lemme 1 : La suite $(r(m))$ possède les propriétés suivantes :

- a) Elle est strictement croissante.
- b) $\forall m : \sup_{1 \leq j \leq t} f(\underline{x}_j^0)/l < r(m)$.
- c) $\lim_{m \rightarrow \infty} r(m) = \sup_{\underline{x} \in R^p} f(\underline{x})$.

Preuve : Les propriétés a) et b) sont triviales. La démonstration de c) est identique à celle qui est donnée dans [1] pages 7-8.

Ainsi, à chaque valeur de $m = 1, 2, \dots$ on peut associer l'ensemble $\{\underline{x} : f(\underline{x}) \geq r(m)\}$, réunion d'un nombre fini d'ensembles disjoints notés $C(r(m),b)$.

On désignera par t_m le nombre de $C(r(m),b)$ associés au seuil $r(m)$, ce qui implique que pour tout m , l'indice b prend ses valeurs dans $\{1, 2, \dots, t_m\}$ avec $1 \leq t_m \leq t$.

Afin de prendre en compte tous les modes lorsque $t > 1$, il convient de choisir correctement la valeur de l qui intervient dans la définition des suites $(r(m))$ et $(r(m,n))$. En effet, il faudra choisir l suffisamment grand pour que :

$$r(0) = \sup_{1 \leq j \leq t} f(\underline{x}_j^0)/l < \inf_{1 \leq j \leq t} f(\underline{x}_j^0) . \text{ (on suppose } r(0) \geq \sigma \text{)}$$

D'autre part, si \underline{x}_j désigne un point situé sur la frontière de A_j tel que $f(\underline{x}_j) = \sup_{\underline{x} \in \text{frontière } A_j} f(\underline{x})$, il faudra choisir l de sorte que :

$$1/l < \inf_{1 \leq j \leq t} (f(\underline{x}_j^0) - f(\underline{x}_j)) .$$

Finalement on choisira l suffisamment grand pour que :

$$1/l < \min \left\{ \inf_{1 \leq j \leq t} f(\underline{x}_j^0) / \sup_{1 \leq j \leq t} f(\underline{x}_j^0) ; \inf_{1 \leq j \leq t} (f(\underline{x}_j^0) - f(\underline{x}_j)) \right\} .$$

Dans ces conditions, on peut être assuré que pour tout j ($j=1,2,\dots,t$), il existera toujours au moins une valeur de m , et donc de $r(m)$, telle que le mode \underline{x}_j^0 n'appartienne qu'à un seul $C(r(m),b)$. Dans la suite, on désignera par $C(r(m),j)$ tout ensemble $C(r(m),.)$ ne contenant que le mode \underline{x}_j^0 .

Lemme 2 : Les ensembles $C(r(m),j)$ sont, pour tout j ,

- a) connexes
- b) disjoints (si $t_m \geq 2$)
- c) bornés et d'intérieur non vide.

Preuve :

a) Cette propriété est une conséquence immédiate de la forme de $f(.)$ autour de ses modes.

b) C'est évident puisque, par définition, chaque $C(r(m),j)$ ne peut contenir qu'un seul mode \underline{x}_j^0 .

c) Par suite des hypothèses faites sur $f(.)$ et sur $r(m)$, pour tout m , $r(m) \geq \sigma$, et $C(r(m),.)$ est borné par sa frontière qui est la projection sur R^p de l'intersection de $f(.)$ avec le plan parallèle au support de $f(.)$, d'ordonnée $r(m)$.

On peut noter que pour tout \underline{x} situé sur la frontière d'un $C(r(m),.)$ on a $f(\underline{x}) = r(m)$, autrement dit, ces frontières sont des courbes iso-densité. Par suite de la continuité et de la forme de $f(.)$, il est évident que toute variation de r conduira à la création d'une famille d'ensembles $C(.,.)$ qui seront soit imbriqués, soit disjoints. Plus précisément, si $r < r'$, on a $\{\underline{x} : f(\underline{x}) \geq r'\} \subset \{\underline{x} : f(\underline{x}) \geq r\}$ donc chaque $C(r',j)$ appartient à un certain $C(r,k)$. Ainsi pour j et k fixés on a :

soit $C(r',j) \subset C(r,k)$, soit $C(r',j) \cap C(r,k) = \emptyset$.

(Hartigan [27] qualifie cette propriété de structure arborescente).

Soit J l'indice j tel que :

$$f(\underline{x}_J^0) = \sup_{\underline{x} \in \mathbb{R}^D} f(\underline{x}) .$$

(Si J n'est pas unique alors le résultat a) du lemme 3 sera vrai pour tous les J).

Lemme 3 : Si $\text{diam } A = \sup \{d(\underline{x}, \underline{x}') ; \underline{x} \text{ et } \underline{x}' \in A\}$ désigne le diamètre de l'ensemble A alors :

- a) $\lim_{m \rightarrow \infty} \text{diam } C(r(m), J) = 0$.
- b) $\forall j \neq J$, il existe m tel que :

$$\text{diam } C(r(m), j) < \text{diam } A_j < +\infty .$$

Preuve :

- a) Considérons l'ensemble $B_\delta = \{\underline{x} : f(\underline{x}) > f(\underline{x}_J^0) - \delta\}$ où $\delta > 0$. Si δ est assez petit on aura :

$$B_\delta \subset A_J \text{ (supposé borné) .}$$

D'autre part, pour tout δ il existe m tel que $C(r(m), J) \subset B_\delta$ (il suffit, en effet, de choisir m tel que $f(\underline{x}_J^0) - \delta < r(m) < f(\underline{x}_J^0)$). Les propriétés de $f(.)$ sur A_J entraînent que l'application $\delta \rightarrow B_\delta$ est décroissante lorsque δ diminue et donc que $\lim_{\delta \rightarrow \infty} B_\delta = \emptyset$ sinon le $\sup_{\underline{x} \in \mathbb{R}^D} f(\underline{x})$

ne serait pas atteint en un singleton \underline{x}_j^0 . Le résultat est alors immédiat.

En résumé, \underline{x}_j^0 étant pour tout m suffisamment grand un point intérieur à $C(r(m), J)$:

$$\lim_{m \rightarrow \infty} C(r(m), J) = \underline{x}_j^0 .$$

b) Le résultat annoncé est une conséquence directe de la façon de choisir l . En effet, $\forall j \neq J$ on a vu qu'il existe m tel que $f(\underline{x}_j) < r(m) < f(\underline{x}_j^0)$ où \underline{x}_j est un point de la frontière de A_j . On en tire le résultat cherché.

Remarque : Le lemme 3 n'assure la convergence des $C(.,.)$ vers \underline{x}_j^0 qu'uniquement pour $j = J$. Néanmoins, on peut, en modifiant légèrement la définition de la suite $(r(m))$ obtenir un lemme équivalent au lemme 3 pour chacun des modes \underline{x}_j^0 . Pour converger vers le mode \underline{x}_j^0 , par exemple, il suffit de définir $r(m)$ par rapport à $f(\underline{x}_j^0)$, c'est-à-dire de poser :

$$r(0) = L f(\underline{x}_j^0) .$$

On peut donc énoncer :

Lemme 4 : Pour tout $j = 1, 2, \dots, t$, à la condition que

$$r(0) = L.f(\underline{x}_j^0) = L \sup_{\underline{x} \in A_j} f(\underline{x}) :$$

$$a) \lim_{m \rightarrow \infty} r(m) = f(\underline{x}_j^0) .$$

$$b) \lim_{m \rightarrow \infty} C(r(m), j) = \underline{x}_j^0 .$$

Preuve : Voir les lemmes 1 et 3.

Lemme 5 : Pour tout n , la suite $(r(m, n))$ est strictement croissante.

Preuve : Evidente.

Lemme 6 : Pour tout n et pour tout $\epsilon_n > 0$, il existe une constante universelle positive D telle que :

$$P\left\{ \sup_{\underline{x} \in \mathbb{R}^p} |F_n(\underline{x}) - F(\underline{x})| > \epsilon_n \right\} \leq D \cdot \exp(-2n \epsilon_n^2) .$$

Preuve : Voir [34] .

Lemme 7 : Si la suite $(S(n))$ remplit les deux conditions suivantes :

- 1) $\lim_{n \rightarrow \infty} S(n) = 0$
- 2) $\lim_{n \rightarrow \infty} n \cdot S_n^{2p} = \infty$

alors $h(n, \underline{x})$ converge presque sûrement vers $f(\underline{x})$.

Preuve : $\forall \underline{x} \in \mathbb{R}^p$ on peut toujours trouver un ensemble d'entiers relatifs q_i , $i = 1, 2, \dots, p$ tel que \underline{x} soit inclus dans l'hypercube ${}^p q_1 \dots q_p$.
D'autre part :

$$S_n^p h(n, \underline{x}) = \Delta^{(1)} \dots \Delta^{(p)} \{F_n[\underline{(q-1)S_n}] - F[\underline{(q-1)S_n}]\} + \Delta^{(1)} \dots \Delta^{(p)} F[\underline{(q-1)S_n}] .$$

En développant $F(\cdot)$ autour de \underline{x} :

$$\Delta^{(1)} \dots \Delta^{(p)} F[\underline{(q-1)S_n}] = S_n^p F(\underline{x}) + o(S_n^{p+1}) . \quad (4)$$

Posons :

$$c_k = \gamma_k (q_k - 1) S_n + (1 - \gamma_k) q_k S_n, \quad k = 1, 2, \dots, p$$

les γ_k prenant indépendamment les uns des autres les valeurs 0 ou 1 .
On a 2^p choix du p -uplet $\gamma = (\gamma_1, \dots, \gamma_p)$, et à chaque choix correspond un $\underline{c}^{(\gamma)} = (c_1^{(\gamma)}, \dots, c_p^{(\gamma)})$ qui est un sommet d'un hypercube, d'où :

$$\Delta^{(1)} \dots \Delta^{(p)} \{F_n(\underline{(q-1)S_n}) - F(\underline{(q-1)S_n})\} = \sum_{\gamma} (-1)^{|\gamma|} (F_n(\underline{c}^{(\gamma)}) - F(\underline{c}^{(\gamma)})) \quad \text{avec } |\gamma| = \sum_k \gamma_k \quad (5)$$

Compte-tenu de (4) et (5) on a :

$$|h(n, \underline{x}) - f(\underline{x})| \leq \varepsilon |F_n(\underline{c}^{(\gamma)}) - F(\underline{c}^{(\gamma)})| / S_n^p + |O(S_n)| \quad (6)$$

Soit φ l'ensemble de définition de $\underline{c}^{(\gamma)}$:

$$\varphi = \{(q_1-1)S_n, q_1S_n, \dots, (q_p-1)S_n, q_pS_n\} .$$

φ étant inclus dans R^p on tire de (6) :

$$\sup_{\underline{x} \in R^p} |h(n, \underline{x}) - f(\underline{x})| \leq 2^p \sup_{\underline{c}^{(\gamma)} \in \varphi} |F_n(\underline{c}^{(\gamma)}) - F(\underline{c}^{(\gamma)})| / S_n^p + |O(S_n)| \leq 2^p \sup_{\underline{x} \in R^p} |F_n(\underline{x}) - \dots - F(\underline{x})| / S_n^p + |O(S_n)| .$$

Donc $\forall \varepsilon > 0$, compte-tenu de la condition (1) sur $(S(n))$, on pourra toujours trouver n suffisamment grand pour que :

$$P\left\{\sup_{\underline{x} \in R^p} |h(n, \underline{x}) - f(\underline{x})| > \varepsilon\right\} \leq P\left\{2^p \sup_{\underline{x} \in R^p} |F_n(\underline{x}) - F(\underline{x})| / S_n^p > \varepsilon/2\right\} .$$

Par suite du lemme 6 :

$$\sum_{n=1}^{\infty} P\left\{\sup_{\underline{x} \in R^p} |h(n, \underline{x}) - f(\underline{x})| > \varepsilon\right\} \leq \sum_{n=1}^{\infty} D \exp(-n\varepsilon^2 S_n^{2p} / 2^{2p+1}) . \quad (7)$$

La condition (2) imposée à la suite $(S(n))$ entraîne la convergence de la série à droite dans l'inégalité (7) et le théorème de Borel-Cantelli complète la démonstration.

Remarques :

1) Si $S(n)$ est de la forme A/n^δ où $A =$ constante alors la condition (2) du lemme 7 entraîne que l'on doit avoir $0 < \delta < 1/2p$.

2) Lorsque le support de $f(\cdot)$ est un pavé de R^p , BERTRAND-RETALI [35] a montré que si $S_n^p = o(\log n/n)$ alors $h(n, \underline{x})$ converge presque complètement vers $f(\underline{x})$.

3) Le cas $p=1$ a été étudié dans [36], [37], [38] et [1].

4) Le lemme 7 peut être complété par les résultats suivants :

a) $E h(n, \underline{x}) = f(\underline{x}) + O(S_n)$.

b) $\text{Var } h(n, \underline{x}) = f(\underline{x})/nS_n^p + O(1/nS_n^{p-1})$.

c) $h(n, \underline{x})$ est asymptotiquement normalement distribuée. Les résultats a) et b) se déduisent aisément de ceux qui précèdent. Le c) est une application du théorème de Moivre-Laplace.

Lemme 8 : Pour tout m , $r(m, n)$ converge presque sûrement vers $r(m)$ lorsque n tend vers l'infini.

Preuve : C'est une conséquence du lemme 7. Voir [1] pour la démonstration.

Pour m donné, posons :

$$\tau_m = \bigcup_{b=1}^{t_m} C(r(m), b) \text{ et}$$

$$\tau_{mn} = \bigcup_{w=1}^{t_{mn}} C(r(m, n), w) .$$

On peut énoncer alors :

Théorème 1 :

Pour tout m , τ_{mn} converge presque sûrement vers τ_m . C'est-à-dire que

$$P (\lim_{n \rightarrow \infty} \tau_{mn} = \tau_m) = 1 .$$

Preuve :

Par définition :

$$\tau_m = \{ \underline{x} : f(\underline{x}) \geq r(m) \}$$

$$\tau_{mn} = \{ \underline{x} : h(n, \underline{x}) \geq r(m, n) \}$$

Les lemmes 7 et 8 assurant de la convergence p.s. de $h(n, \underline{x})$ vers $f(\underline{x})$ et de $r(m, n)$ vers $r(m)$ le résultat annoncé s'en déduit immédiatement.

Remarques :

1) A la limite l'ensemble τ_{mn} peut donc différer de l'ensemble τ_m d'un ensemble de mesure nulle. C'est donc un résultat moins fort que celui donné par SAGER dans [18] (L'écart maximum entre les ensembles τ_{mn} et τ_m converge p.s. vers zéro). Sur le plan pratique cette différence ne présente pas d'inconvénient. En outre, on bénéficiera ici du fait que, contrairement à SAGER, les régions modales n'ont pas besoin d'être convexes. SAGER lui-même avait d'ailleurs souligné la difficulté de mise en oeuvre de son estimateur à cause de l'absence d'algorithme efficace de recherche du plus petit ensemble convexe contenant un nombre donné de points.

2) Le théorème 1 a pour corollaire que :

$$\forall m : t_{mn} \xrightarrow{\text{p.s.}} t_m .$$

Si on s'intéresse à l'estimation du (ou des) mode(s) de $f(\cdot)$ on pourra appliquer l'algorithme aux observations situées dans chacune des régions modales $C(r(m,n),w)$ comme on l'a déjà souligné à la fin du second paragraphe. Soit \underline{x}_{wn}^0 un point intérieur de $C(r(m,n),w)$ alors :

Théorème 2 :

Lorsque m et n tendent vers l'infini :

$$\underline{x}_{wn}^0 \xrightarrow{\text{p.s.}} \text{un des modes } \underline{x}_j^0 \text{ de } f(\cdot) .$$

Preuve :

C'est une conséquence immédiate du théorème 1 et du lemme 4 (en restreignant le domaine de définition de \underline{x} à $C(r(m,n),\underline{x}) \subset \mathbb{R}^p$).

Comme précédemment on ne pourra pas savoir vers quel mode particulier converge \underline{x}_{wn}^0 . Tout ce que l'on pourra dire c'est que pour n et m suffisamment grands \underline{x}_{wn}^0 sera proche de l'une des modes \underline{x}_j^0 sans pouvoir préciser la valeur exacte de j .

Notons enfin que si $t = 1$ (f unimodale) il n'y a évidemment plus d'ambiguïté pour les théorèmes 1 et 2.

4 - UN EXEMPLE D'APPLICATION A UN ECHANTILLON GAUSSIEN BIDIMENSIONNEL

Afin de pouvoir être utilisé plus facilement, l'algorithme décrit plus haut, a été aménagé comme indiqué ci-après.

L'histogramme servant de base à l'estimateur est celui des effectifs plutôt que des densités de fréquences (il y a une différence d'un facteur d'échelle constant, égal à $n.S_n^p$, entre les deux représentations).

D'autre part, pour accélérer le processus d'évolution des régions modales avec l'accroissement du seuil $r(m,n)$, on a fixé à un le pas de variation de $r(m,n)$ au lieu de $1/l$.

Pour finir, on arrête le processus d'estimation lorsque $r(m,n)$ devient supérieur à $\sup_{\underline{x} \in R^p} h(n, \underline{x})$.

Ces modifications font que les étapes 4 et 5 de l'algorithme de définition de $r(m,n)$ sont maintenant :

4) Si $\sup_{\underline{x} \in R^p} h(n, \underline{x}) - r(m-1, n) > 1$: aller en 5 ,

Si non : fin .

5) Calculer $r(m, n) = r(m-1, n) + 1$.

Pour illustrer le comportement de cet algorithme on l'a appliqué à un échantillon artificiel constitué de $n = 300$ observations tirées de façon équiprobable des trois lois normales bivariées suivantes :

Vecteurs des espérances
transposés

(8 8)
(14 14)
(14 2)

Matrice des variances -
covariances

4 I_2
4 I_2
4 I_2

où I_2 est la matrice identité d'ordre deux. Cet échantillon est représenté sur la figure 1.

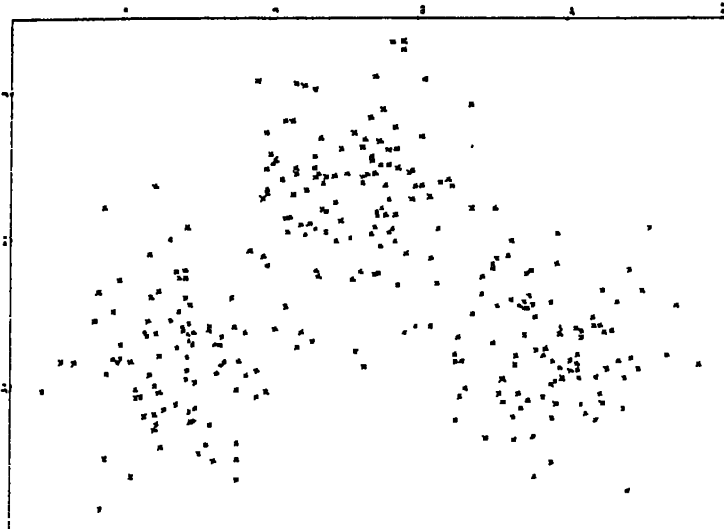


fig. 1 - échantillon

Puisque l'on ne fixe aucune contrainte de forme pour les régions modales en dehors de leur connexité (simple ou multiple), il en résulte que deux carrés (de côté S_n) de R^2 appartiennent à une même région modale dès qu'ils ont au moins un point commun. Ainsi, la figure 2 représente deux exemples possibles de régions modales (on suppose, bien sûr, que $h(n, \underline{x}) \geq r(m, n)$ dans tout carré hachuré).

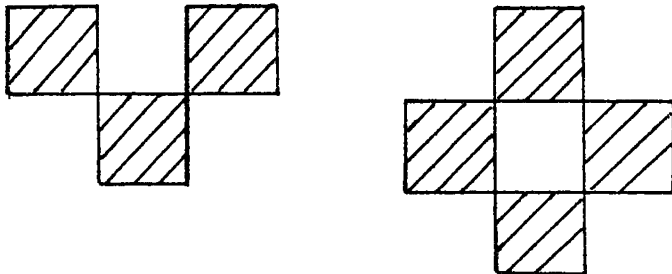


fig. 2 - Deux exemples de région modale (hachurée)

L'algorithme a été utilisé avec les paramètres suivants :

$$l = 5 \quad \text{et} \quad S(n) = A/n^{\frac{1}{2p+1}} \quad \text{où} \quad p = 2 \quad \text{et} \quad A = \text{constante.}$$

Les tableaux 1, 2 et 3 ci-dessous indiquent le nombre de régions modales obtenues pour quelques valeurs de A et du seuil d'effectif r(m,n) .

r(m,n)	2,6	3,6	4,6	5,6	6,6	7,6	8,6	9,6	10,6	11,6	
A = 5	2	4	3	3	3	3	4	3	3	2	(r(0,n) = $\frac{13}{5} = 2,6$)

r(m,n)	3	4	5	6	7	8	9	10	11	12	13	14	
A = 6	1	1	2	2	3	3	3	3	3	3	2	2	(r(0,n) = $\frac{15}{5} = 3$)

r(m,n)	3,6	4,6	5,6	6,6	7,6	8,6	9,6	10,6	11,6	12,6	13,6	14,6	15,6	16,6	17,6	
A = 7	1	1	1	1	2	3	3	3	3	3	3	3	3	1	1	(r(0,n) = $\frac{18}{5} = 3,6$)

Tableaux 1, 2, 3 - Nombre de régions modales

On constate que, pour chaque valeur de A , le nombre de régions modales obtenu reste constant et égal à trois pour une large gamme de valeurs de r(m,n) . Lorsqu'elle se présente, (c'est le cas en particulier lorsque les densités des modes ont des valeurs voisines) cette constance du nombre de modes pourra donc être utilisée pour déterminer le nombre de modes réellement présents dans l'échantillon ainsi que pour choisir la valeur de A (ou de S_n) .

Pour une valeur donnée de S_n (ou de A) , désignons par N₃ le maximum du nombre de fois où l'on trouve trois régions modales successives lorsque r(m,n) augmente. Sur l'ensemble des simulations réalisées (pour des raisons de place on ne peut fournir tous ces résultats ici) on observe que, lorsque A augmente (et donc S_n) , N₃ commence par croître puis diminue pour les grandes valeurs de A . On peut donc mettre en évidence une ou plusieurs valeurs de S_n conduisant à de grandes valeurs de N₃ .

Ici la valeur $S_n = 2,237$ (correspondant à $A = 7$), pour laquelle on obtient une plage de $N_3 = 8$ valeurs successives égales à trois, pourra être considérée comme voisins de l'optimum. Pour tous les résultats qui sont donnés dans la suite on conservera donc cette valeur de A .

On constate également que de faibles valeurs de S_n associées à de petites valeurs de $r(m,n)$ conduisent à un nombre excessif de régions modales tandis qu'un découpage trop grossier de R^2 ne permet de détecter qu'un seul mode quel que soit $r(m,n)$.

Le choix de l ne paraît pas critique dans la mesure où il suffit d'adopter une valeur assez grande afin de partir d'un $r(0,n)$ pas trop grand et donc de "n'oublier" aucun mode. On voit sur le tableau 3, par exemple, qu'il faudrait prendre une valeur de l égale à 1 pour "rater" les 3 modes au cours de la croissance de $r(m,n)$.

Sur un autre plan, l'algorithme proposé permet facilement d'effectuer une estimation de la mesure de Lebesgue des régions modales obtenues. Pour cela, il suffit de compter le nombre d'hypercubes (ici des carrés) définissant la région modale considérée puis de multiplier ce nombre par la mesure S_n^p (ici S_n^2) d'un hypercube.

Dans le cas de l'échantillon traité, on peut comparer ces estimations aux valeurs théoriques. En effet, la densité de la population échantillonnée est ici :

$$f(x_1, x_2) = \frac{1}{3} \sum_{i=1}^3 f_i(x_1, x_2)$$

où $f_i(x_1, x_2) = \frac{1}{8\pi} \exp \left(-\frac{1}{2} \left(\left(\frac{x_1 - \mu_{i1}}{2} \right)^2 + \left(\frac{x_2 - \mu_{i2}}{2} \right)^2 \right) \right)$

avec	8	si	$i = 1$
$\mu_{i1} =$	14	si	$i = 2, 3$
	8	si	$i = 1$
$\mu_{i2} =$	14	si	$i = 2$
	2	si	$i = 3$

Dans ces conditions, un plan parallèle au plan $x_1 O x_2$ d'ordonnée r , coupe $f(x_1, x_2)$ suivant trois ellipses coplanaires. d'équations respectives :

$$\frac{(x_1 - \mu_{i1})^2 + (x_2 - \mu_{i2})^2}{8 \text{Log}(n S_n^2 / 24\pi r)} = 1, \quad i = 1, 2, 3.$$

Chacune de ces ellipses ayant une même aire donnée par :

$$8 \pi \text{Log}(n S_n^2 / 24\pi r).$$

Pour $A = 7$, soit $S_n = 2,237$, l'algorithme donne par exemple, suivant la valeur du seuil $r(m, n)$:

. $r = 9,6$:

1 région modale d'aire égale à 25,02

2 régions modales d'aire respective égale à 15,01 .

Soit une aire moyenne observée de 18,35 alors que la valeur théorique est 17,31 .

. $r = 12,6$:

1 région modale d'aire égale à 20,02

1 région modale d'aire égale à 15,01

1 région modale d'aire égale à 10,01

Soit une aire moyenne estimée de 15,01 pour une valeur théorique de 10,71 .

. $r = 15,6$:

3 régions modales d'aire égale à 5,00 chacune.

Soit une aire moyenne estimée de 5,00 pour une valeur théorique de 5,50 .

Ces résultats montrent, qu'en présence de régions modales théoriques de même mesure, la moyenne des mesures fournies par l'algorithme constitue une bonne estimation. On observe aussi que la variance expérimentale des

mesures individuelles diminue lorsque le seuil $r(m,n)$ augmente. Ceci est certainement dû au fait qu'en augmentant $r(m,n)$ on accroît la séparation spatiale des modes (au moins pour l'exemple étudié).

Pour continuer, on a noté dans le tableau 4, l'évolution du nombre de points contenu dans les régions modales en fonction du seuil $r(m,n)$:

$r(m,n)$	5,6	6,6	7,6	8,6	9,6	10,6	11,6	12,6	13,6	14,6	15,6	16,6
Nombre de régions	1	1	2	3	3	3	3	3	3	3	3	1
Nombre de points par région	241	225	50 140	71 53 50	71 44 41	71 44 31	60 44 31	60 44 31	60 31 18	46 31 18	16 16 18	18

Tableau 4 - Effectifs des régions modales.

Contrairement à ce que l'on pourrait attendre, on constate une certaine dissymétrie entre les effectifs des régions sauf pour la valeur 15,6 du seuil. Une analyse visuelle détaillée de l'échantillon montre que ce déséquilibre est réel (la dispersion des points du nuage étant différente suivant le mode considéré) et l'algorithme ne fait donc que mettre en évidence ce fait dû aux hasards de l'échantillonnage. Pour la valeur 15,6 du seuil, l'aire des régions modales est plus faible et on constate, en conséquence, une atténuation des écarts comparativement à ceux observés sur des régions de plus grande étendue.

Si on s'intéresse à l'estimation des modes de la population, il faut se donner la valeur du seuil d'effectif n_S . Les résultats qui suivent (tableaux 5 et 6) illustrent le comportement de l'algorithme pour deux valeurs de n_S :

$$. n_S = 40$$

$r(m,n)$	9,6	10,6	11,6	12,6	13,6	14,6	15,6	16,6
nombre de modes estimés	0	1	1	1	2	2	3	1

Tableau 5 - Nombre de modes estimés pour $n_S = 40$.

$$n_5 = 20$$

$r(m,n)$	9,6	10,6	11,6	12,6	13,6	14,6	15,6	16,6
nombre de modes estimés	0	0	0	0	1	1	3	1

Tableau 6 - Nombre de modes estimés pour $n_5 = 20$

Pour $n_5 = 20$ et $r(m,n) = 15,6$ les positions des modes estimés (centres de gravité des régions modales) sont :

$$\begin{array}{lll} x_1 = 7,83 & x_1 = 14,54 & x_1 = 14,54 \\ x_2 = 5,54 & x_2 = 1,12 & x_2 = 14,54 \end{array}$$

alors que les modes théoriques sont :

$$\begin{array}{lll} x_1 = 8 & x_1 = 14 & x_1 = 14 \\ x_2 = 8 & x_2 = 2 & x_2 = 14 \end{array}$$

Enfin mentionnons que sur l'ordinateur NAS du C.I.R.C.E. il faut moins de 4 centièmes de seconde de temps C.P.U pour effectuer le traitement correspondant à une valeur du seuil $r(m,n)$ pour un échantillon de taille $n = 300$.

5 - CONCLUSION

Les performances de l'algorithme étudié sont conditionnées en pratique par le choix de l et par celui de S_n . Suite à la remarque faite à la fin du lemme 1, le choix de l peut être facilité en considérant qu'il suffit de choisir une valeur suffisamment grande pour éviter "l'oubli" de certains modes. On pourrait même rendre l'algorithme autonome sur ce point en notant que, pour un histogramme d'effectifs, le seuil de départ $r(0,n)$ peut être pris égal à un de sorte qu'il suffirait de poser :

$$l = \sup_x h(n, \underline{x}) .$$

Le choix de S_n (ou de A) est plus délicat mais, comme le montre l'exemple traité au paragraphe quatre, il pourra reposer sur l'observation du nombre de régions modales associées à différents seuils $r(m,n)$. En effet, si l'on porte dans un tableau le nombre de modes obtenus en fonction de S_n et de $r(m,n)$ on observe une zone, orientée approximativement suivant la diagonale principale du tableau, pour laquelle le nombre de régions modales reste constant. Il suffira alors de choisir une valeur de S_n correspondant à l'amplitude maximum de cette zone. D'autres études (voir [24] et [28] par exemple) ont abouti à un résultat similaire.

Notons, pour finir, que cet algorithme utilise l'estimateur de la densité le plus simple qui soit : un histogramme ordinaire. On pourrait le remplacer par un histogramme généralisé (estimateur à noyau) et bénéficier ainsi des propriétés de ce type d'estimateur (la continuité en particulier).

BIBLIOGRAPHIE

- 1 J.P. ASSELIN de BEAUVILLE : "Un algorithme d'estimation du mode
Etude de la convergence".
Pub. Inst. Stat. Univ. Paris, 1979, XXIV, fasc. 3-4,
p. 1-29.
- 2 E. PARZEN : "On estimation of a probability density function and
mode".
Ann. Math. Statist. (1962), 33, p. 1065-1076.
- 3 H. CHERNOFF : "Estimation of the mode".
Ann. Inst. Statist. Math. (1964), 16, p. 31-41.
- 4 E.A. NADARAYA : "On non parametric estimates of density function
and regression curves".
Theor. Probability Appl. (1965), 10, p. 186-190.
- 5 T. DALENIUS : "The mode. A neglected statistical parameter".
J. Roy. Statist. Soc. (1965), Ser. A, 128, p. 110-111.
- 6 U. GRENANDER : "Some direct estimates of the mode".
Ann. Math. Statist. (1965), 36, p. 131-138.
- 7 J.H. VENTER : "On estimation of the mode".
Ann. Math. Statist. (1967), 38, p. 1446-1455.
- 8 B.P. LIENTZ : "Results on non parametric modal intervale".
SIAM Journal on Applied Math. (1970), 19, p. 355-366.
- 9 E.J. WEGMAN : "A note on the estimation of the mode".
Ann. Math. Statist. (1971), 42, 6, p. 1909-1915.
- 10 H. EKBLÖM : "A Monte-Carlo investigation of mode estimators in
samples".
Appl. Statist. (1972), 21, 2, p. 177-184.
- 11 J. FRITZ : "Stochastic approximation for finding local maxima of
probability densities".
Studia Scientiarum Mathematicarum Hungarica (1973),
p. 309-322.
- 12 T. ROBERTSON et J.D. CRYER : "An iterative procedure for estimating
the mode".
J.A.S.A. (1974), 69, 348, p. 1012-1016.

- 13 T.W. SAGER : "Consistency in non parametric estimation of the mode".
Ann. Statist. (1975), 3, p. 698-706.
- 14 W.F. EDDY : "Optimum kernel estimators of the mode".
Ann. Statist. (1980), 8, 4, p. 870-882.
- 15 M. SAMANTA : "Nonparametric estimation of the mode of a multivariate density".
S. Afr. Statist. J. (1973), 7, p. 109-117.
- 16 V.D. KONAROV : "On the asymptotic normality of the mode of multidimensional distributions".
Theor. Probability Appl. (1974), 19, p. 794-799.
- 17 T.W. SAGER : "Estimation of a multivariate mode".
Ann. Statist. (1978), 6, p. 802-812.
- 18 T.W. SAGER : "An iterative method for estimating a multivariate mode and isopleth".
J.A.S.A. (1979), 74, 366, p. 329-339.
- 19 T. CACOULLOS : "Estimation of a multivariate density".
Ann. Inst. Statist. Math. (1966), 13, p. 179-189.
- 20 D. WISHART : "Mode analysis : A generalisation of nearest neighbour which reduces chaining effects".
Numerical Taxonomy (1969), ed. A.J. COLE, Academic Press, London, p. 282-311.
- 21 D. WISHART : "Numerical classification method for deriving natural classes". Nature (1969), 221, 4, p. 97-98.
- 22 I. GITMAN et M.D. LEVINE : "An algorithm for detecting unimodal fuzzy sets and its application as a clustering technique".
IEEE Trans. Comput. (1970), C-19, p. 583-593.
- 23 I. GITMAN : "An algorithm for nonsupervised pattern classification".
IEEE trans. system man, cybern. (1973), SMC-3, 1, p. 66-74.
- 24 D.J. EIGEN, F.R. FROMM, R.A. NORTHOUSE : "Cluster analysis based on dimensional information with applications to feature selection and classification".
IEEE trans. system. man, cybern. (1974), SMC-4, 3.
- 25 R. MIZOGUCHI et M. SHIMURA : "Nonparametric learning without a teacher based on mode estimation".
IEEE Trans. Comput. (1976), C-25, 11.
- 26 J.A. HARTIGAN : "Clustering algorithms".
(1975), Ed. Wiley.

- 27 J.A. HARTIGAN : "Clusters as modes".
First International Symposium on data analysis and informatics (1977), IRIA, p. 433-448.
- 28 C. VASSEUR et J. POSTAIRE : "Convexité des fonctions de densité : Application à la détection des modes en reconnaissance des formes".
RAIRO Automatique (1979), 13, 2, p. 171-188.
- 29 C. VASSEUR et J. POSTAIRE : "A convexity testing method for cluster analysis".
IEEE Trans. Syst. Man. Cybern. (1980), SMC-10, 3.
- 30 R.S. CHHIKARA et D.T. REGISTER : "A numerical classification method for partitioning of large multidimensional mixed data set".
Technometrics, (1979), 21, 4, p. 531-537.
- 31 R.M. CORMACK : "A review of classification".
J. Roy. Statist. Soc., (1971), A, 134, p. 321-367.
- 32 H.H. BOCK : "Clustering by density estimation".
Analyse des Données et Informatique (1979), INRIA, p. 173-186.
- 33 M. LIMOURI et J. POSTAIRE : "Utilisation des histogrammes unidimensionnels en classification automatique".
Reconnaissance des formes et intelligence artificielle (1979), AFCET-IRIA, p. 373-380.
- 34 J. KIEFER et J. WOLFOVITZ : "On the deviations of the empirical distribution function of vector chance variables".
Trans. Am. Math. Soc. (1958), 87, p. 173-186.
- 35 M. BERTRANS-RETALI : "Convergence uniforme stochastique d'un estimateur d'une densité de probabilité dans R^S ".
C.R. Acad. Sc. (1974), A, 278, p. 451-453.
- 36 P. REVESZ : "Testing of density functions".
Periods. Math. Hungar. (1971), 1, p. 35-44.
- 37 P. REVESZ : "On empirical density functions".
Periods. Math. Hungar. (1972), 2, p. 85-110.
- 38 P. REVESZ : "A strong law of the empirical density function".
Transactions sixth Prague Conf. (1973), Prague, p. 747-753.