

La classification automatique par les statistiques d'ordre

Cluster analysis by order statistics

par El Hassan SBAI

Ecole Supérieure de Technologie,
Université Moulay Ismaïl
Route d'Agouray, B.P. 3103
50006, Toulal – Meknès, Maroc.

résumé et mots clés

Le problème de la classification automatique non supervisée peut être abordé en détectant les modes de la fonction de densité de probabilité sous-jacente à la distribution des observations disponibles. Dans cet article, on montre comment ces modes peuvent être mis en évidence par des transformations adaptatives basées sur les statistiques d'ordre. Des résultats expérimentaux sont présentés en dernier lieu pour montrer la robustesse de ces transformations de statistiques d'ordre.

Fonction de densité de probabilité (fdp), statistiques d'ordre, classification automatique, mode, gradient.

abstract and key words

The problem of unsupervised pattern classification can be tackled by detecting the modes of the underlying probability density function of the data. In this paper, we show how the modes can be detected by adaptive transformations based on order statistics. Finally, experimental results are presented to show the robustness of these transformations.

Probability density function, order statistics, clustering, mode, gradient.

1. introduction

La classification automatique joue un rôle important en reconnaissance des formes, plus particulièrement lorsqu'aucune information n'est connue préalablement. Dans le contexte de la classification non supervisée, les méthodes de classification essaient de partitionner un ensemble d'observations en les groupant en un nombre de classes statistiques. Dans ce qui suit, nous verrons qu'une classe peut être définie comme un ensemble d'observations dont la densité est supérieure à celle du volume environnant.

De nombreuses méthodes de classification automatique statistique se ramènent à des procédures de détection des modes de la fonction de densité de probabilité sous-jacente à la distribution des données disponibles [GIT 73, DEV 82, ASS 89]. Les techniques de détection des modes les plus connues sont celles où les modes peuvent être considérés comme des maxima locaux de la fdp.

Comme toutes les approches basées sur l'utilisation d'opérateurs différentiels, ces méthodes sont connues pour être sensibles aux irrégularités locales de la distribution des observations et tendent à générer de nombreux modes parasites [KIT 76]. Une alternative consiste à assimiler les modes de la fdp à des régions de l'espace où cette fonction est localement convexe [VAS 80]. Cette approche améliore considérablement la robustesse de détection des modes par rapport aux techniques faisant appel aux notions du gradient. Malgré ses avantages apparents, il est important de noter que les domaines d'observations peuvent être à cheval entre les régions concaves et convexes et que les suppositions pour le test de convexité sont violées. En plus, il est noté que ce test est sensible aux irrégularités de la distribution des observations. Des techniques d'étiquetage probabiliste itératif, ou relaxation, ont été développées pour améliorer la robustesse de cette technique [POS 94]. Une autre approche du problème de la détection des modes a été abordée par localisation de leurs contours en utilisant les opérateurs différentiels multidimensionnels [TOU 89]. Là aussi,

le même auteur a développé une technique d'étiquetage probabiliste permettant ainsi de diminuer l'effet des irrégularités de la distribution des données [TOU 88]. Ces cinq dernières années et loin des considérations statistiques, une méthode d'analyse locale de la distribution des observations utilisant des critères géométriques relevant de la morphologie mathématique binaire a été proposée [POS 93]. Les mêmes auteurs ont développé une technique qui combine les concepts de la morphologie mathématique multinationaux aux propriétés de convexité de la fdp tout en cherchant à améliorer la discrimination entre les modes [ZHA 94].

Le but de cet article est d'introduire les concepts de statistiques d'ordre dans le domaine de la classification automatique statistique. Il s'agit de soumettre la fdp à des transformations adaptatives introduisant l'opérateur gradient pour extraire ses modes. Là aussi, on cherche à améliorer davantage la séparation entre les modes lorsque le chevauchement entre les classes est important.

Le reste de cet article est organisé comme suite : dans la deuxième section, nous rappelons la discrétisation de la fdp sous-jacente à la distribution des observations multidimensionnelles ; la troisième et quatrième section rappellent brièvement l'opérateur différentiel de Prewitt et les filtres de statistiques d'ordre respectivement. Ces derniers sont utilisés pour déterminer la $i^{\text{ème}}$ statistique d'ordre d'un ensemble de n éléments ; dans la cinquième section, nous proposons un filtre de statistiques d'ordre adaptatif. Dans la sixième section nous présenterons quelques résultats expérimentaux.

2. discrétisation de la fonction de densité de probabilité

Les opérateurs de détection des modes qui seront introduits dans le présent article sont définis sur une version discrète de la fonction de densité de probabilité. Cette fonction est estimée à partir des observations disponibles selon un algorithme rapide qui assure de bonnes performances à la procédure de classification de données.

Supposons que l'on dispose d'un échantillon $\Gamma = \{X_1, X_2, \dots, X_q, \dots, X_Q\}$ de Q observations indépendantes et identiquement distribuées suivant une fonction de densité de probabilité $f(x)$, pour estimer cette fonction lorsque l'analyste ne dispose d'aucune autre information que celle extraite des observations disponibles, ce dernier doit faire appel à des techniques d'estimation non paramétriques [WEG 72]. La solution la plus connue à ce problème est la méthode du noyau, initialement proposée par Rosenblatt [ROS 56] dans le cas monovarié, puis approfondie

par Parzen [PAR 62] et étendue au cas multidimensionnel par Cacoullos [CAC 65] et Murthy [MUR 66].

Dans un espace de dimension N , l'estimateur est de la forme :

$$\hat{f}(X) = \frac{1}{Q} \sum_{q=1}^Q h_Q^{-N} \Phi \left(\frac{X - X_q}{h_Q} \right)$$

où $\Phi(X)$ est une fonction réelle positive appelée noyau de largeur h_Q . Les opérateurs de statistiques d'ordre dont l'application nécessite une discrétisation de l'espace de représentation des données, nous amène à opter pour le noyau cubique de Parzen pour estimer la fonction de densité de probabilité. Ce dernier est donné par :

$$h_Q^{-N} \Phi \left(\frac{X}{h_Q} \right) = \begin{cases} \left(\frac{1}{2h_Q} \right)^N & \text{si } \frac{|X_n|}{h_Q} \leq 1, n = 1, 2, \dots, N \\ 0 & \text{autrement} \end{cases}$$

$$\text{où } X = [x_1, x_2, \dots, x_n, \dots, x_N]^T.$$

Le choix de ce noyau est également motivé par des considérations de programmation sur ordinateur numérique. Cependant, malgré ses avantages apparents, l'utilisation de cette technique d'estimation de la fdp nécessite certaines précautions pour éviter le temps de calcul prohibitif lorsque la dimension des données croît. Un algorithme rapide a été proposé par Postaire et Vasseur [POS 82] pour pallier cet inconvénient.

L'implantation de la procédure d'estimation de la fdp par la méthode du noyau nécessite une discrétisation de l'espace \mathcal{R}^N de représentation des données. Une transformation linéaire de l'espace permet de normaliser la plage de variation de chacune des composantes $x_{q,1}, x_{q,2}, \dots, x_{q,n}, \dots, x_{q,N}$ des observations $X_Q, q = 1, 2, \dots, Q$ telle que : $\max_q x_{q,n} - \min_q x_{q,n} = h_Q, N$ $n = 1, 2, \dots, N$.

Chaque axe de l'espace de données est ainsi partitionné en R intervalles adjacents et égaux. Cette discrétisation définit un réseau de R^N hypercubes de côté unité. L'algorithme rapide estime la fdp uniquement aux centres des hypercubes non vides de ce réseau hypercubique dont le nombre Q' ne peut dépasser celui des observations disponibles. De plus, cette procédure permet de connaître le nombre d'observations situées dans chacun des hypercubes non vides, donc la valeur de l'estimateur en son centre. Ainsi, en désignant par \underline{X} l'ensemble discret constitué des hypercubes non vides du réseau, le traitement des données se limite aux centres $P_r, r = 1, 2, \dots, Q', Q' \leq Q$ des hypercubes, de coordonnées à valeurs entières, constituant cet ensemble.

A la fin de cette phase de discrétisation, toute l'information disponible pour assurer la classification des observations est contenue dans l'ensemble discret de valeurs estimées de la fdp aux points d'échantillonnage.

3. l'opérateur différentiel de Prewitt

L'opérateur différentiel de Prewitt permet d'estimer le gradient de la fdp en tout point du réseau d'échantillonnage [TOU 89]. Le module du gradient $G(P_r)$ au point $P_r (r = 1, 2, \dots, R)$ peut être estimé sous la forme :

$$\hat{G}(P_r) = \sum_{i=1}^N |a_i|$$

où les coefficients a_i sont donnés par l'expression :

$$a_i = \frac{\sum_{V_m(P_r)} p_{r',i} \hat{f}(p_{r',1}, p_{r',2}, \dots, p_{r',N})}{\sum_{V_m(P_r)} (p_{r',i})^2}, i = 1, 2, \dots, N$$

avec $\sum_{V_m(P_r)} (p_{r',i})^2 = 2(1^2 + 2^2 + \dots + m^2)(2m+1)^{N-1}$ et $V_m(P_r)$

étant le voisinage hypercubique, de taille m , du point P_r défini par les points $P_{r'}$ tel que :

$$V_m(P_r) = \{p_{r',1}, \dots, p_{r',i}, \dots, p_{r',N} / |p_{r',j} - m < p_{r',j} < p_{r',j} + m ; i = 1, 2, \dots, N\}.$$

Notons que les coefficients a_i ont été obtenus en supposant que le point P_r est l'origine du voisinage V_m . Dans ce qui suit on prend $m = 1$.

4. les filtres de statistiques d'ordre (SO) pour les ensembles et les fonctions

Dans cette section seront rappelées brièvement les idées de base des filtres de statistiques d'ordre qui seront utilisées dans la définition des détecteurs des modes. Pour plus de détails, le lecteur intéressé pourra consulter par exemple les références [DAV 81, MAR 85, MAR 87, HAR 91]. Ces filtres constituent une classe de filtres discrets non linéaires connus pour leurs propriétés robustes de filtrage du bruit dans le domaine du traitement du signal, de la parole et des images. Les filtres de statistiques d'ordre peuvent être appliqués à tout signal discret de dimension finie.

Soit $S = \{X_1, X_2, \dots, X_n\}$ un ensemble de n nombres réels. Supposons qu'on trie ces n nombres dans l'ordre croissant de telle sorte que $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$; le $i^{\text{ème}}$ nombre de cette liste triée $X_{(i)}$ est appelé la $i^{\text{ème}}$ statistique d'ordre de l'ensemble fini $S, i = 1, 2, \dots, n$. Les exemples de statistiques d'ordre les plus connues sont : les extrêmes $X_{(n)}$ et $X_{(1)}$, l'étendue $W = X_{(n)} - X_{(1)}$, la déviation à partir de la moyenne $X_{(n)} - \bar{X}$ et la médiane $X_{(n/2)}$. L'idée de base des filtres de statistiques d'ordre consiste à remplacer le point courant par une statistique de la séquence correspondant aux éléments du voisinage considéré. Les valeurs maximales et minimales de la séquence triée sont les opérations primitives (dilatation et érosion) pour tous les filtres morphologiques et le voisinage correspond à l'élément structurant. En plus, les filtres de statistiques d'ordre sont des transformations morphologiques des signaux par des ensembles parce qu'ils satisfont les 4 principes morphologiques [SER 82].

Soit V_1 un voisinage composé de k éléments. La $i^{\text{ème}}$ statistique d'ordre de la fonction f (f définie sur Z^N, Z ensemble des entiers et N un entier positif) par V_1 est la fonction

$$[SO^i(f : V_1)](x) = i^{\text{ème}} SO \text{ de } \{f(y) :$$

$$y \in V_{1x}\}, x \in Z^N; 1 \leq i \leq n \text{ et } V_{1x} = \{x + a : a \in V_1\}$$

est le translaté de V_1 au point x .

L'érosion (dilatation) de la fonction f par V_1 en tout point x est obtenue en translatant V_1 au point x et en prenant le minimum (maximum) de f dans V_1 :

$$(f \ominus V_1^s)(x) = \text{Inf}_{y \in V_{1x}} \{f(y)\} \left((f \oplus V_1^s)(x) = \text{Sup}_{y \in V_{1x}} \{f(y)\} \right).$$

V_1^s est le symétrique de V_1 par rapport à l'origine :

$$V_1^s = \{-\nu : \nu \in V_1\}.$$

Dans notre cas V_1 est le même que V_1^s .

En combinant la définition de statistique d'ordre avec les définitions d'érosion et de dilatation, on voit que la première ($i = 1$) statistique d'ordre de toute fonction par un voisinage V_1 coïncide avec son érodé par V_1 . De même, la $n^{\text{ème}}$ statistique d'ordre est égale au dilaté par V_1 . Ainsi, les érosions et les dilatations sont des cas particuliers des filtres de statistiques d'ordre. Dans la suite, nous nous intéressons aux filtres de statistiques d'ordre des ensembles puisque la fonction de densité de probabilité est représentée par un ensemble discret de valeurs estimées de cette fonction.

Dans la section suivante nous développerons un filtre adaptatif de détection des modes de la fonction de densité de probabilité basé sur les statistiques d'ordre.

5. détection des modes de la fonction de densité de probabilité par les statistiques d'ordre

5.1. procédure de détection des modes

Dans de nombreux cas en classification automatique statistique, les classes peuvent être considérées comme des régions à forte densité séparées par des régions à faible densité. Pour cela, les classes peuvent être identifiées en recherchant les régions à forte densité, appelées modes, dans l'espace des données.

Zhang et Postaire [ZHA 94] ont proposé d'accomplir la dilatation morphologique lorsque la fonction de densité de probabilité sous-jacente est localement concave, *i.e.* lorsque le rapport des densités moyennes estimées $\frac{\hat{\rho}_1}{\hat{\rho}_2}$ dans deux voisinages homothétiques est supérieur à 1, et d'accomplir l'érosion morphologique de cette fonction lorsqu'elle est convexe, *i.e.* lorsque le rapport est inférieur à 1. Dans cet article, nous proposons de considérer les transformations morphologiques en terme de procédure de filtrage de statistiques d'ordre. L'érosion consiste à prendre comme sortie du filtre la première valeur de la liste des valeurs de la fdp estimée triées dans l'ordre croissant. Cette valeur est la plus petite dans le voisinage hypercubique considéré. De même, la dilatation consiste à prendre comme sortie du filtre de statistiques d'ordre la dernière valeur de cette liste, qui est la plus grande dans le voisinage hypercubique.

Avec le voisinage hypercubique utilisé par Zhang et Postaire, composé de 3^N points d'échantillonnage, le rang de sortie du filtre morphologique dépendant des propriétés de convexité de la fdp est égal à 1 lorsque le rapport $\frac{\hat{\rho}_1}{\hat{\rho}_2} < 1$ (érosion), et ce rang est égal à 3^N lorsque $\frac{\hat{\rho}_1}{\hat{\rho}_2} > 1$ (dilatation). La figure 1(a) représente cette fonction binaire.

Au lieu d'utiliser seulement les valeurs maximales et minimales de la fdp, *i.e.* le premier et le dernier élément de la liste du filtre de statistiques d'ordre, nous proposons de prendre tous les éléments de la liste, en utilisant une relation spécifique entre les valeurs prises par le gradient de la fdp et les rangs des valeurs ordonnées de la fonction de densité dans le voisinage considéré. Le rang de sortie du filtre de statistiques d'ordre est d'autant plus grand que le gradient de la fdp est grand.

Les statistiques d'ordre les plus élevées sont utiles pour réhausser les modes de la fonction de densité de probabilité alors que les

statistiques d'ordre les plus faibles sont utiles pour creuser les vallées entre les modes de cette fonction. Donc, au lieu d'appliquer systématiquement le même filtre de statistiques d'ordre à tout l'échantillon à analyser, les statistiques d'ordre les plus élevées sont appliquées aux points d'échantillonnage où la fonction de densité possède un gradient élevé, alors que les statistiques d'ordre les plus faibles sont appliquées aux points d'échantillonnage où la fonction de densité possède un faible gradient.

Dans le processus de détection des modes, le filtre proposé changera en chaque point du réseau d'échantillonnage, d'où l'utilisation du terme adaptatif. La transformation non linéaire accomplie en chaque point du réseau est déterminée par les variations locales de la fonction de densité de probabilité en ce point.

Considérons l'ensemble discret du réseau hypercubique précédemment défini (section II), $\underline{X} = \{P_r; r = 1, 2, \dots, Q'\}$, où Q' est le nombre d'hypercubes non vides. Considérons aussi une fenêtre hypercubique V_1 (fenêtre $3^*3^*...^*3$) qui englobe k points, k est un nombre impair ($k = 3^N$), N est la dimension de l'espace des données. A chaque point P_r de coordonnées entières $P_{r,1}, P_{r,2}, \dots, P_{r,N}$, du réseau de discrétisation, les k points d'échantillonnage qu'englobe la fenêtre hypercubique V_1 sont : $P_{r_1}, P_{r_2}, \dots, P_{r_k}$. Les valeurs de la fonction de densité de probabilité estimée de ces k points de la fenêtre sont :

$$\hat{f}(P_{r'}) ; P_{r'} \in V_1(P_r) = \{P_{r_i}, i = 1, 2, \dots, k\}$$

où $V_1(P_r)$ est la fenêtre hypercubique centré au point P_r .

On définit le vecteur ordonné $\overset{\circ}{Y}(P_r) = \text{ordre}[\hat{f}(P_{r'}) ; P_{r'} \in V_1(P_r)] = [\hat{f}_{(1)}, \hat{f}_{(2)}, \dots, \hat{f}_{(k)}]$ où $\hat{f}_{(1)} \leq \hat{f}_{(2)} \leq \dots \leq \hat{f}_{(k)}$. Le rang d'ordre g ou la $g^{\text{ème}}$ statistique d'ordre du vecteur ordonné $\overset{\circ}{Y}(P_r)$ est défini par

$$g(\overset{\circ}{Y}(P_r)) = g(\hat{f}_{(1)}, \hat{f}_{(2)}, \dots, \hat{f}_{(k)}) =$$

$$Y_g(P_r) : Y_g(P_r) \in \{\hat{f}(P_{r'}) : P_{r'} \in V_1(P_r)\} \quad 1 \leq g \leq k.$$

La valeur minimale des k éléments, par exemple, correspond à la première statistique d'ordre ($g = 1$). De même, le maximum correspond à la $k^{\text{ème}}$ statistique d'ordre ($g = k$).

Dans la fenêtre spécifiée, la sortie du filtre au point P_r est une constante contrainte à être une statistique d'ordre du vecteur $\overset{\circ}{Y}(P_r)$. Le filtre introduit l'opérateur différentiel pour sélectionner un rang parmi ceux des points d'échantillonnage qu'englobe la fenêtre V_1 . La fonction utilisée pour sélectionner le rang du filtre

$$\text{est de la forme } 1 + \left[(1 - \exp \left(- \left(\frac{\hat{G}}{a} \right)^l \right) * 3^N - 0.5 \right], \text{ où } [x]$$

est la partie entière de x , \hat{G} est le module du gradient de Prewitt [TOU 89] et les paramètres a et l sont des entiers. Notons que le rang prend une valeur entre 1 et 3^N . D'autres fonctions peuvent être adoptées pour choisir le rang du filtre.

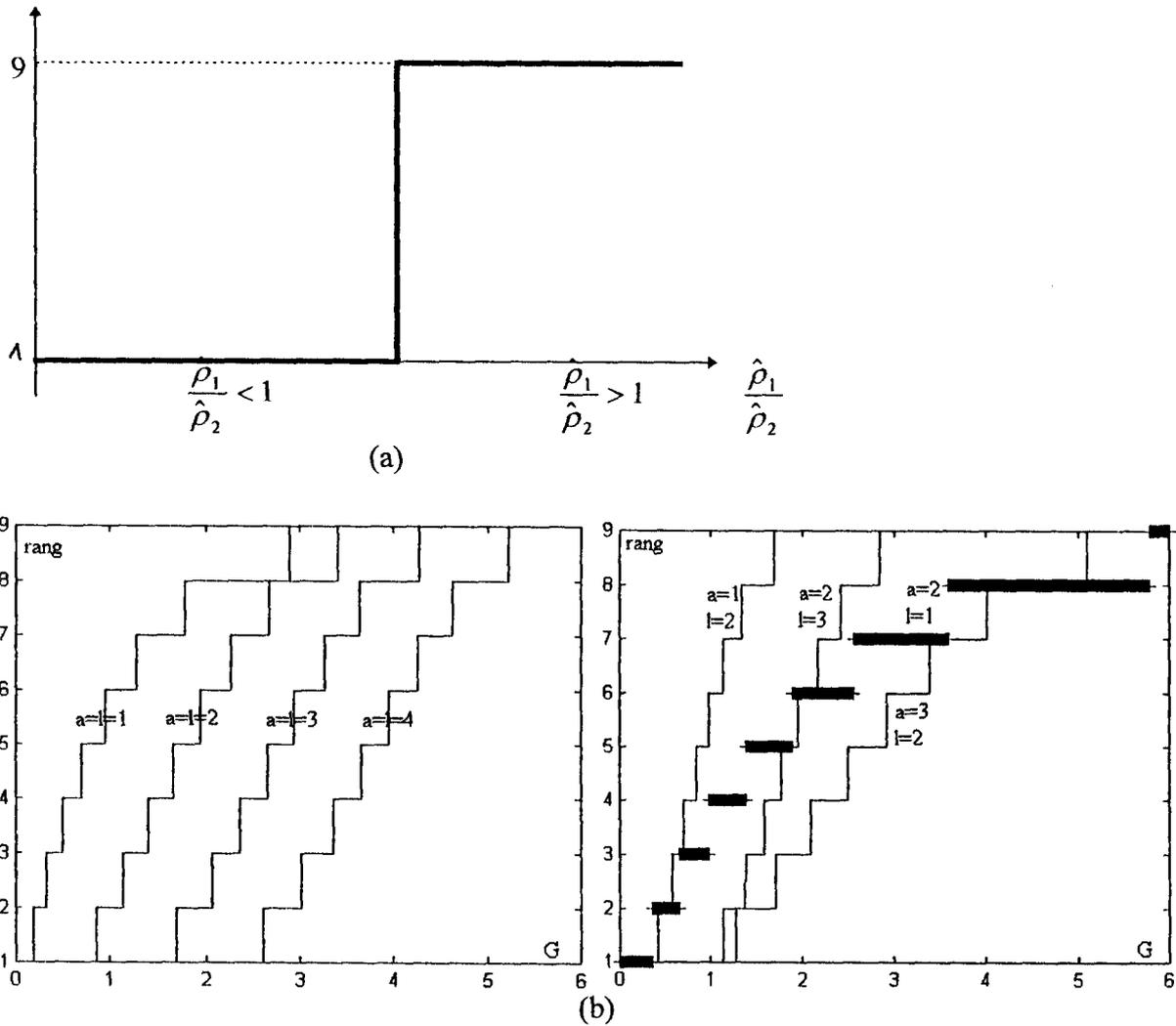


Figure 1. – Tracé du rang du filtre : a) rang du filtre en fonction du rapport des densités moyennes estimées; b) rang du filtre en fonction du module du gradient.

Par exemple :

$$\bullet \Psi_1(\hat{G}) = 1 + \left[\left(\frac{(\hat{G}/a)^l}{1 + (\hat{G}/a)^l} \right) * 3^N - 0.5 \right];$$

$$\bullet \Psi_2(\hat{G}) = 1 + \left[\left(\frac{\exp(\frac{\hat{G}}{a})^l - 1}{\exp(\frac{\hat{G}}{a})^l + 1} \right) * 3^N - 0.5 \right];$$

$$\bullet \Psi_3(\hat{G}) = 1 + \left[\left(\frac{1}{1 + \exp(-[(\frac{\hat{G}}{a})^l + \Theta])} \right) * 3^N - 0.5 \right],$$

Θ : paramètre.

Ces fonctions donnent perceptiblement des résultats similaires.

Le choix d'un faible rang à partir des rangs des points d'échantillonnage qu'englobe la fenêtre V_1 est utile pour décroître les valeurs de la fonction de densité de probabilité dans les vallées, alors que le choix d'un rang élevé est utile pour accentuer les valeurs de la fdp dans les régions modales. Par conséquent, au lieu de prendre le même rang de sortie et d'appliquer le même opérateur de statistiques d'ordre à la fdp, un rang élevé est sélectionné aux points d'échantillonnage où le gradient de la fdp est important, alors qu'un rang faible est sélectionné aux points où le gradient de cette fonction est faible. Cette stratégie d'analyse de la fdp est en fait basée sur les variations locales de celle-ci. Le choix arbitraire de la formule $1 + \left[\left(1 - \exp(-(\frac{\hat{G}}{a})^l) \right) * 3^N - 0.5 \right]$, parmi d'autres, réside dans la détermination d'un rang adéquat qui per-

met en fin de compte de détecter facilement les modes de la fdp. La nature non linéaire des filtres de statistiques d'ordre rend leur analyse très difficile. Ainsi, l'adoption d'un filtre de statistiques d'ordre approprié en vue de la détection des modes de la fdp est en relation étroite avec le choix de la formule et des paramètres a et l .

Notons que le gradient de la fonction de densité de probabilité au voisinage des maxima locaux est faible. Ainsi donc, le filtre de statistiques d'ordre ne permet pas d'augmenter l'amplitude de la fonction de densité au voisinage des maxima locaux de celle-ci. La sélection d'un faible rang dans les régions du réseau d'échantillonnage au voisinage des maxima locaux n'a pas d'effet néfaste sur la détection des modes puisque les valeurs de la fdp aux points d'échantillonnage qu'englobe la fenêtre hypercubique V_l sont non nulles.

Le tracé du rang en fonction du module du gradient de la fonction de densité de probabilité pour différentes valeurs de a et l est donné sur la figure 1(b). Les effets de variation des paramètres a et l aident à la compréhension de l'adaptation du rang du filtre de statistiques d'ordre et donnent un aperçu sur le choix de leurs valeurs. Le graphe 1(b) montre bien l'intérêt de tenir compte de a et l comme paramètres de réglage de l'algorithme de détection des modes. En effet, si par exemple $(a = 2, l = 3)$, $(a = 3, l = 2)$, $(a = 3, l = 3)$, $(a = 3, l = 4)$ ou $(a = 4, l = 4)$, la sortie du filtre est égale à zéro si les estimées de la fdp du premier rang de la liste jusqu'au rang sélectionné sont égales à zéro, même pour des gradients non proches de zéro. Si par exemple $(a = 1, l = 1)$, $(a = 1, l = 2)$ ou $(a = 2, l = 1)$ moins de sorties nulles du filtre apparaissent. En plus, ces sorties nulles du filtre n'apparaissent que pour des gradients proches de zéro, ou si plusieurs estimées de la fdp du voisinage considéré sont nulles. La stratégie de filtrage proposée fournit donc plus de points d'échantillonnage à valeurs nulles lorsque les paramètres a et l prennent des valeurs élevées, ce qui permet une meilleure séparation des classes surtout lorsque le chevauchement est important.

Le comportement du filtre de statistique d'ordre adaptatif dépend donc en plus du pas de discrétisation de l'espace des données, des paramètres a et l . Dans ce qui suit et en vue de creuser davantage les vallées de la fdp, on prend $a = l = 3$.

5.2. ajustement de l'algorithme de détection des modes

Le comportement de la procédure de détection des modes basée sur les statistiques d'ordre dépend largement de l'ajustement de la résolution h_Q du processus de discrétisation. Si h_Q est très petit, tous les hypercubes du réseau d'échantillonnage seront connexes et les petits modes ne seront pas détectés. Si h_Q est très grand, la procédure tendra à générer un grand nombre de petits modes non significatifs. En fait, le pas de discrétisation dépend non seulement du nombre total Q d'observations de l'échantillon et de la dimension N de l'espace de représentation des données, mais

aussi de la structure de la distribution de ces données. Lorsque rien n'est connu *a priori* à propos de cette structure, l'ajustement de ce paramètre est basé sur le concept de stabilité du nombre de modes [EIG 74]. L'ajustement de la résolution h_Q consiste donc à analyser les résultats obtenus pour une grande marge de variation de ce paramètre. Il a été prouvé que le choix de ce paramètre au milieu du plus grand intervalle à l'intérieur duquel le nombre de modes détectés reste constant et différent de 1 est une bonne procédure pour optimiser un nombre important d'algorithmes de classification [TOU 88].

6. résultats expérimentaux

Les performances des filtres de statistiques d'ordre, dans le cadre de leur application à la détection des modes de la fonction de densité de probabilité, sont analysées en utilisant des données générées artificiellement. Pour montrer l'intérêt de la procédure proposée, les résultats obtenus sont comparés à ceux résultant des transformations morphologiques dépendant des propriétés de convexité de la fonction de densité [ZHA 94] et à ceux obtenus en analysant la convexité de cette fonction [VAS 80]. Cette dernière méthode consiste à estimer les régions modales où la densité est convexe permettant ainsi de définir des surfaces de séparation pour les différentes classes.

Exemple 1 :

Le premier exemple, représenté sur la figure 2(a), est constitué de trois classes gaussiennes équiprobables de 600 données bidimensionnelles (cf. Tableau 1). Le graphe de la fdp estimée est visualisé sur la figure 2(b). Les figures 2(c) et 2(d) montrent l'effet du paramètre de discrétisation h_Q sur le nombre de modes détectés respectivement par les opérateurs de statistiques d'ordre et l'approche de la morphologie mathématique introduisant les propriétés de convexité de la fdp [ZHA 94]. Rappelons que l'ajustement du pas de discrétisation est basé sur le concept de stabilité du nombre de modes détectés [EIG 74]. Lorsque l'ensemble des données à analyser est composé de plusieurs classes, les modes correspondant doivent apparaître pour un grand intervalle de valeurs de h_Q .

D'après les figures 2(c) et 2(d), on remarque que les résultats obtenus par l'algorithme proposé sont nettement meilleurs que ceux obtenus par la méthode utilisée par Zhang et Postaire [ZHA 94]. En effet, la plage de stabilité du nombre de modes détectés par les statistiques d'ordre suivant les variations de h_Q s'étend entre 18 et 38, alors que la plage de stabilité obtenue en utilisant la deuxième technique est réduite à l'intervalle [16-27].

Les résultats de la figure 2(e) sont obtenus en analysant la convexité de la fdp. Cette approche est basée sur le test qui détermine localement la convexité de la fonction de densité,

Données générées				Résultats obtenus		
Distribution	Vecteur moyenne	Matrice de covariance	Probabilité à priori	Vecteur moyenne	Matrice de covariance	Probabilité à priori
1	$X_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\Sigma_1 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$	0.33	$X_1 = \begin{bmatrix} -0,02 \\ -0,25 \end{bmatrix}$	$\Sigma_1 = \begin{bmatrix} 2,96 & -0,19 \\ -0,19 & 2,28 \end{bmatrix}$	0,318
2	$X_2 = \begin{bmatrix} 0 \\ 7 \end{bmatrix}$	$\Sigma_2 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$	0,33	$X_2 = \begin{bmatrix} -0,13 \\ 6,99 \end{bmatrix}$	$\Sigma_2 = \begin{bmatrix} 3,01 & -0,09 \\ -0,09 & 2,38 \end{bmatrix}$	0,332
3	$X_3 = \begin{bmatrix} 7 \\ 4 \end{bmatrix}$	$\Sigma_3 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$	0.33	$X_3 = \begin{bmatrix} 7,23 \\ 4,18 \end{bmatrix}$	$\Sigma_3 = \begin{bmatrix} 3,13 & 0,29 \\ 0,29 & 2,84 \end{bmatrix}$	0,350

Tableau 1. – Paramètres statistiques de l'exemple 1.

directement à partir des observations disponibles. Dans ce premier exemple où la taille de l'échantillon est importante, la procédure de test de la convexité de la densité de probabilité donne des résultats favorables.

La figure 2(f) présente les trois modes détectés par le filtre de statistique d'ordre, avec $h_Q = 28$ qui correspond au milieu de la plus grande plage de valeurs pour lesquelles le nombre de modes détectés reste constant.

Le résultat de la classification par affectation de chaque observation à la classe associée au mode le plus proche est présenté sur la fig. 2(g). Le taux d'erreur de classification, estimé comme le rapport du nombre d'observations mal classées au nombre total d'observations disponibles, est de 3.1 %. Ce taux peut être comparé au taux d'erreur théorique minimum de Bayes qui est égal à 1.8 %.

Exemple 2 :

Un autre exemple de données bidimensionnelles est présenté (cf. Fig. 3(a)) pour évaluer les performances de la méthode proposée. L'échantillon est constitué de trois classes gaussiennes équiprobables de 150 observations chacune. Les paramètres statistiques sont :

$$X_1 = \begin{bmatrix} 0 \\ 3 \end{bmatrix}, X_2 = \begin{bmatrix} 3 \\ 0 \end{bmatrix}, X_3 = \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \Sigma_1 = \Sigma_2 = \Sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

La figure 3(b) montre les 3 modes détectés pour un pas de discrétisation $h_Q = 25$ qui correspond au milieu de la plus grande plage de valeurs pour lesquelles le nombre de modes détectés reste constant ([20-29]). Comme dans l'exemple précédent, les résultats obtenus montrent la capacité de la méthode proposée à détecter les modes de la fonction de densité de probabilité dans le cas des échantillons de petite taille. Le taux d'erreur associé à cette procédure est de 9.7 % alors que le taux d'erreur théorique

minimum de Bayes est de 7.3 %. La plus grande plage de stabilité du nombre de modes détectés par les transformations morphologiques dépendant des propriétés de convexité de la fonction de densité de probabilité s'étend entre 14 et 18 seulement. Le plus grand intervalle de stabilité du nombre de modes détectés par le test de convexité de la densité de probabilité s'étend entre 20 et 26. Quoique cette dernière méthode constitue une alternative intéressante pour la classification des données multidimensionnelles, elle tend à générer des erreurs dans le processus d'étiquetage des hypercubes convexes dans le cas des échantillons de petite taille. Les résultats estimés sont :

$$\begin{aligned} \hat{X}_1 &= \begin{bmatrix} -0.19 \\ 3.06 \end{bmatrix}, \hat{X}_2 = \begin{bmatrix} 3.8 \\ -0.3 \end{bmatrix}, \hat{X}_3 = \begin{bmatrix} 2.88 \\ 2.99 \end{bmatrix} \\ \hat{\Sigma}_1 &= \begin{bmatrix} 0.66 & 0.13 \\ 0.13 & 0.92 \end{bmatrix}, \hat{\Sigma}_2 = \begin{bmatrix} 0.82 & -0.002 \\ -0.002 & 0.74 \end{bmatrix}, \\ \hat{\Sigma}_3 &= \begin{bmatrix} 0.94 & 0.22 \\ 0.22 & 1.02 \end{bmatrix} \end{aligned}$$

$p_1 = 0.382, p_2 = 0.313, p_3 = 0.304, p_i, i = 1, 2, 3,$ sont les probabilités *a priori* estimées.

Exemple 3 :

Dans le but d'explorer la structure des données de l'ensemble à analyser, sans aucune information *a priori*, les opérateurs de statistiques d'ordre permettent la mise en évidence des groupements de forme et de taille très variées. Le troisième exemple, fig. 4(a), utilisé pour évaluer les performances de l'algorithme, est constitué de 200 données bidimensionnelles non gaussiennes. Les coordonnées de chaque observation sont définies par :

$$\begin{aligned} x_1 &= A_1 \cos\Theta + B_1 \\ x_2 &= A_1 \sin\Theta + B_2 \end{aligned}$$

La classification automatique par les statistiques d'ordre

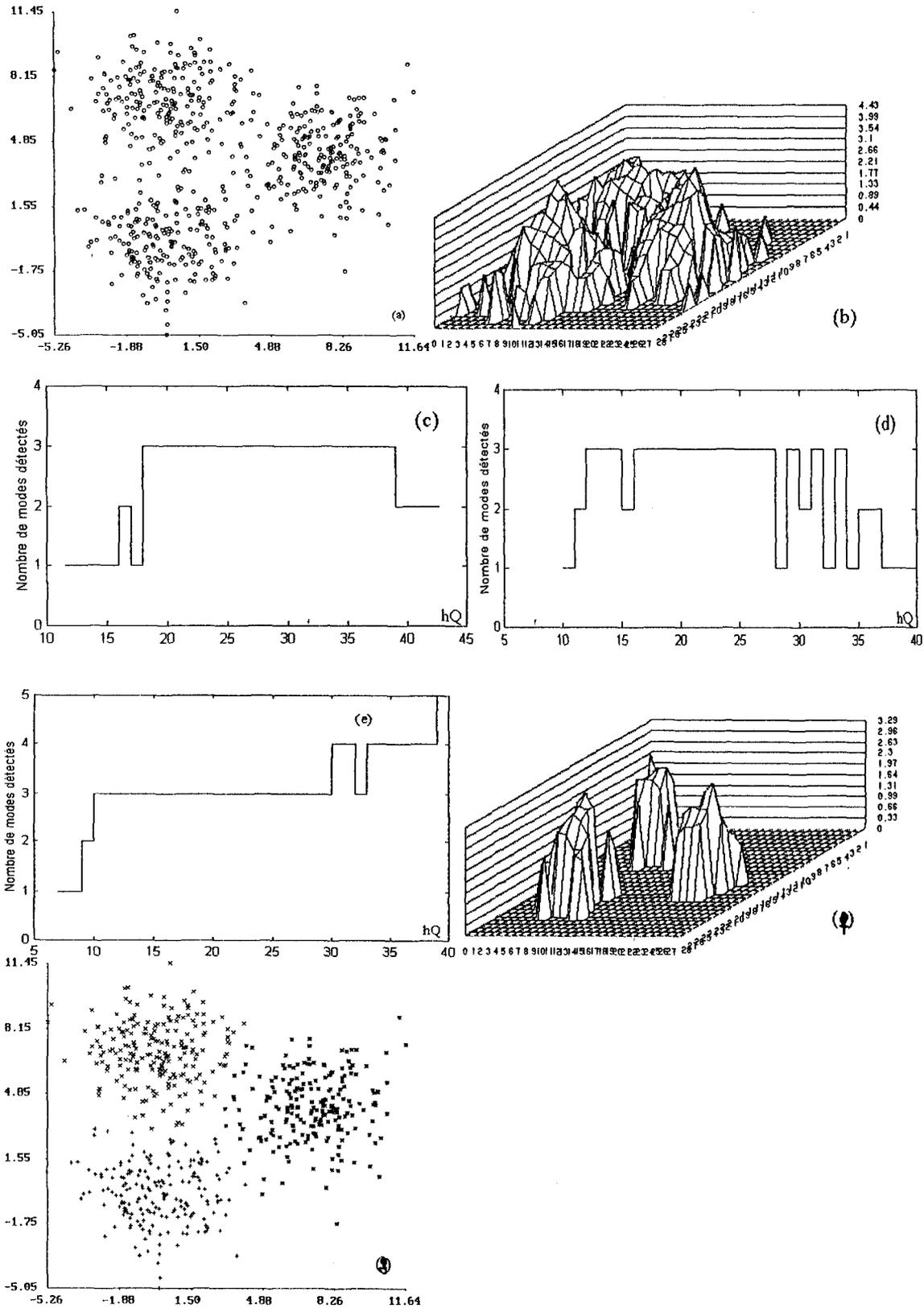


Figure 2. – Détection des modes classification des observations de l'exemple 1 : a) Données à classer ; b) Estimation de la fdp ; c, d, e) Effet de la variation du pas de discrétisation sur le nombre de modes détectés respectivement par application d'opérateurs de statistiques d'ordre, de la morphologie mathématique introduisant les propriétés de convexité et du test de convexité de la fdp ; f) Les modes détectés ; g) Données classées.

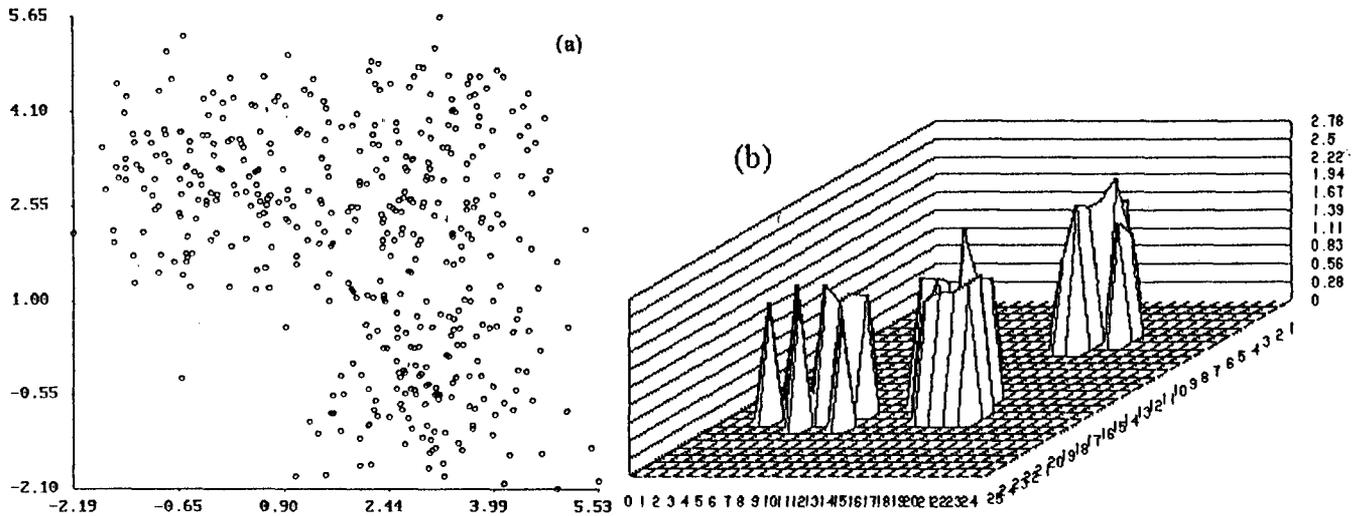


Figure 3. – Détection des modes de l'exemple 2. a) ensemble des données ; b) Les 3 modes détectés.

Distribution	Θ (deg)	B_1	B_2	A_1	A_2
1	$m = 0$ $S = 45$	$\mu_1 = 0$ $\sigma_1 = 8$	$\mu_2 = 0$ $\sigma_2 = 8$	15	15
2	$m = 180$ $s = 45$	$\mu_1 = 0$ $\sigma_1 = 8$	$\mu_2 = -16$ $\sigma_2 = 8$	15	15

Tableau 2. – Paramètres des distributions de l'exemple 2

où θ est une variable aléatoire normale de moyenne m et de variance s et où B_1 et B_2 sont des variables aléatoires normales de moyennes μ et de variance σ [FUK 84]. Les paramètres de l'échantillon considéré sont consignés dans le tableau 2.

La figure 4(b) montre l'estimateur brut de la fdp. Le domaine de stabilité du nombre de modes détectés suivant les variations du paramètre de discrétisation h_Q est illustré par la fig. 4(c). Les deux modes détectés par la procédure proposée sont présentés sur la fig. 4(d) avec une résolution $h_Q = 18$ qui correspond au milieu du plus grand intervalle de valeurs pour lesquelles le nombre de modes détectés reste constant.

Le résultat de la classification est présenté sur la fig. 4(e). Le taux d'erreur de classification est de 3 % alors que le taux d'erreur obtenu par l'algorithme classique des k -moyennes [MAC 67] est de 4 %. C'est en fait pour des données non gaussiennes que réside l'intérêt principal de l'approche proposée qui tient compte de la forme des classes détectées.

Les figures 4(f) et 4(g) montrent l'effet du paramètre h_Q sur le nombre de modes détectés respectivement par les transformations morphologiques dépendant de l'analyse de convexité de la fdp et par le test de convexité de cette fonction respectivement. Ces

résultats montrent bien la faiblesse de la première procédure à détecter les modes de la fdp. Dans cet exemple, les modes sont bien définis par les propriétés de convexité pour une résolution comprise entre 7 et 21. Ce domaine est moins étendu que celui obtenu par le filtre de statistiques d'ordre proposé.

Exemple 4 :

Les trois exemples présentés précédemment ont concerné plus particulièrement la classification automatique bidimensionnelle. Le quatrième exemple présenté fig. 5a), dont les caractéristiques sont consignées dans le tableau 3, concerne les données tridimensionnelles issues de deux classes gaussiennes. Pour la visualisation des résultats de la procédure de détection des modes, une série de plans parallèles et perpendiculaires à l'axe ox_3 est utilisée (fig. 5b). Les groupements obtenus après classification sont représentés sur la figure 5c).

Pour cet ensemble des données, le taux d'erreur de classification est de 1.3 % alors que le taux d'erreur théorique minimum est égale à 1.1 %. La différence entre le taux d'erreur actuel et sa valeur optimale correspond à seulement 2 observations mal classées sur un total de 1000.

7. conclusion

Dans cet article, les opérateurs de statistiques d'ordre ont été appliqués à l'analyse de la fonction de densité de probabilité sous-jacente à la distribution des observations disponibles. Il a été montré comment ces opérateurs tendent à réhausser les modes de la fonction de densité de probabilité et creuser les vallées entre eux. Les filtres de statistiques d'ordre adaptatifs

La classification automatique par les statistiques d'ordre

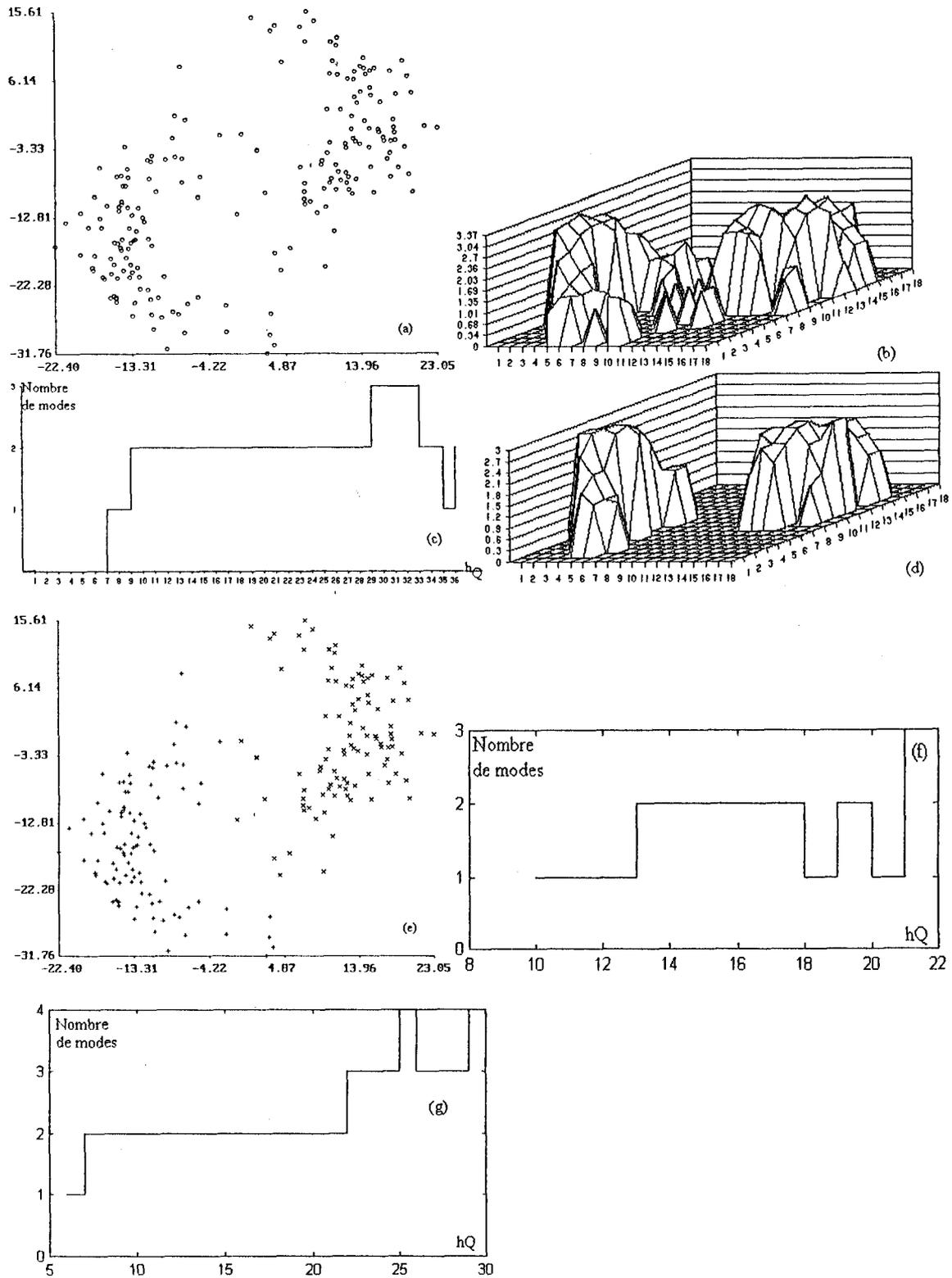


Figure 4. – Détection des modes et classification des données de l'exemple 3 : a) Observations à classer ; b) Estimation de la fdp ; c) Effet de la variation du pas de discrétisation h_Q sur le nombre de modes détectés par application de la procédure proposée ; d) Les modes détectés ; e) Observations classées ; f, g) Effet de la variation de h_Q sur le nombre de modes détectés resp. par application de la morphologie mathématique dépendant des propriétés de convexité de la fdp et du test de convexité de la fonction de densité.

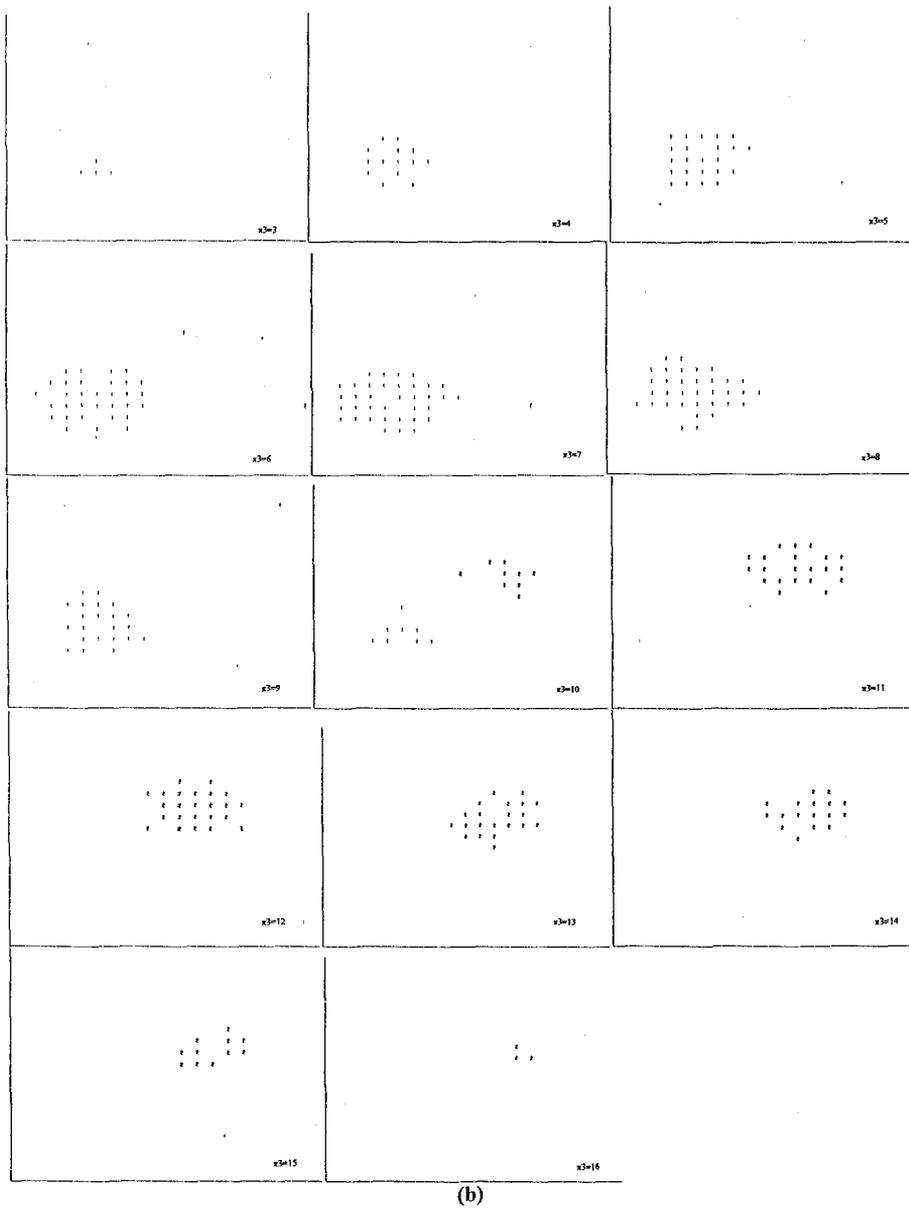
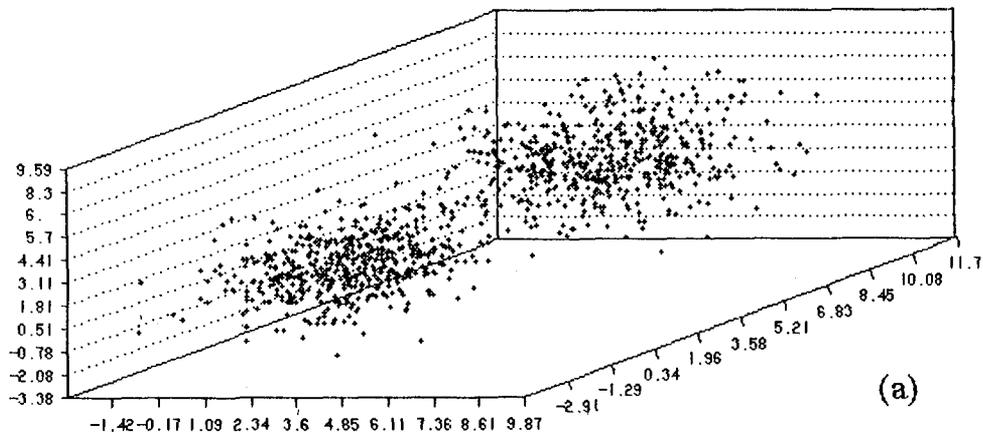


Figure 5.

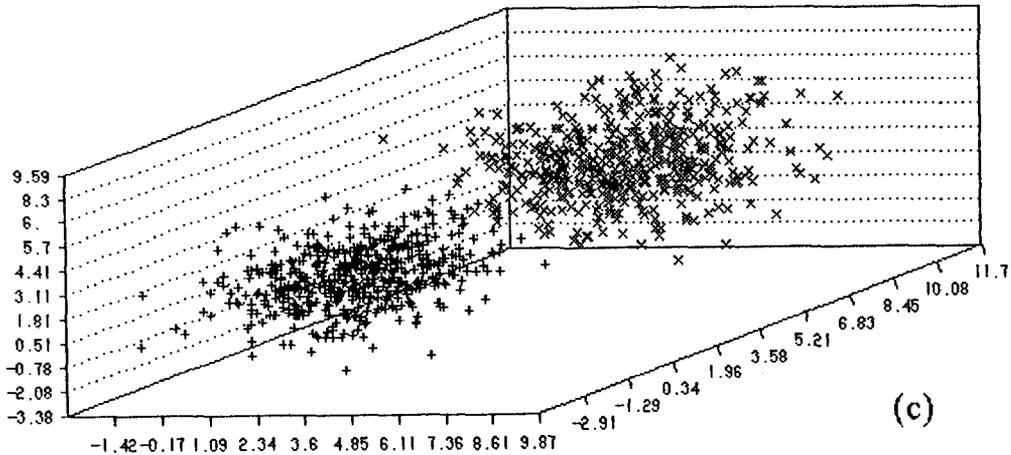


Figure 5 (suite). – Détection des modes et classification des données de l'exemple 4 : a) Observations à classer ; b) Les modes détectés dans une série de 14 plans parallèles et perpendiculaires à l'axe ; c) Observations classées.

Distribution	Vecteur moyenne	Matrice de covariance	Probabilité à priori
1	$M_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ $\widehat{M}_1 = \begin{bmatrix} 0,98 \\ 1,02 \\ 1,07 \end{bmatrix}$	$\Sigma_1 = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ $\widehat{\Sigma}_1 = \begin{bmatrix} 1,93 & 0,15 & 0,14 \\ 0,15 & 1,99 & 0,1 \\ 0,14 & 0,1 & 1,89 \end{bmatrix}$	$P_1 = 0,5$ $\widehat{P}_1 = 0,5$
2	$M_2 = \begin{bmatrix} 5 \\ 5 \\ \frac{5}{cr} \end{bmatrix}$ $\widehat{M}_2 = \begin{bmatrix} 5,14 \\ 5,05 \\ 4,94 \end{bmatrix}$	$\Sigma_2 = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ $\widehat{\Sigma}_2 = \begin{bmatrix} 2,78 & -0,12 & -0,09 \\ -0,12 & 2,51 & -0,08 \\ -0,09 & -0,08 & 2,99 \end{bmatrix}$	$P_2 = 0,5$ $\widehat{P}_2 = 0,5$

Tableau 3. – Paramètres statistiques de l'exemple 3.

qui permettent la mise en évidence des classes en présence, tiennent compte des variations locales de la fdp. Les résultats expérimentaux présentés et les performances, mesurées par le taux d'erreur, ainsi obtenues montrent bien l'intérêt des concepts de statistiques d'ordre en classification automatique statistique. Malgré ses avantages apparents, l'approche proposée présente des limitations au niveau de son application dans des espaces de dimension élevée qui ne présentent pas un grand nombre d'observations.

La classification finale de l'ensemble des données est accomplie par affectation de chaque observation à la classe associée au mode le plus proche.

BIBLIOGRAPHIE

[ASS 89] J.P. Asselin de Beauville, Panorama sur l'utilisation du mode en classification automatique, *APII*, vol. 23, n°2, pp. 113-137, 1989.
 [CAC 66] T. Cacoullos, Estimation of a multivariate density, *Ann. Inst. Statist Math.*, vol. 18, pp. 178-189, 1966.
 [DAV 81] H.A. David, *Order Statistics*, New York : Wiley, 1981.

[DEV 82] A. Devijver, J. Kittler, *Pattern recognition : a statistical approach*, Englewood cliffs, NJ, Prentice-Hall International, 1982.
 [EIG 74] D. G. Eigen, F. R. Fromm and R. A. Northouse, Cluster analysis based on dimensional information with applications to feature selection and classification, *IEEE Trans. Sys. Man Cybern.*, vol. SMC-4, pp. 284-294, 1974.
 [FUK 84] K. Fukunaga, J.M. Mantock, Nonparametric data reduction, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, pp. 115-118, 1984.
 [GIT 73] I. Gittman, An algorithm for nonsupervised pattern classification, *IEEE Trans. Sys. Man Cybern.*, vol. SMC-3, n° 1, pp. 66-74, 1973.
 [HAR 91] R.C. Hardie, G.R. Arce, Ranking in R_p and its use in multivariate image estimation, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 1, n°2, pp. 197-209, June 1991.
 [KIT 76] J. Kittler, A locally sensitive method for cluster analysis. *Pattern Recognition*, Pergamon Press, vol. 8, pp. 23-33, 1976.
 [MAC 67] J. Macqueen, Some methods for classification and analysis of multivariate observations, in Proc. 5th Symp. Math. Stat. Prob., pp. 281-297, 1967.
 [MAR 85] P. Maragos, R.W. Schafer, A unification of linear, median, order statistics, and morphological filters under mathematical morphology, in Proc. 1985 *IEEE ICASSP*, Tampa, FL, Mar. 1985, pp. 34.8.1-34.8.4.

- [MAR 87] P. Maragos, R.W. Schafer, Morphological filters – Part II : Their relations to median, order-statistics, and stack filters. *IEEE Trans. On Acoustics Speech, and Signal Processing*, vol. ASSP-35, n°8, 1987.
- [MUR 66] V. K. Murthy, Nonparametric estimation of multivariate densities with applications. Multivariate analysis, academic press, New York, pp. 43-56, 1966.
- [PAR 62] E. Parzen, On estimation of a probability density function and modes, *Ann. Math. Statist.*, vol. 33, pp. 1065-1076, 1962.
- [POS 82] J.-G. Postaire, C.-P.-A. Vasseur, A fast algorithm for non parametric probability density estimation, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-4 n°6, pp. 663-666, 1982.
- [POS 93] J.-G. Postaire, R.D. Zhang and C. Lecocq-Botte, Cluster analysis by binary morphology, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-15, n°2, pp. 170-180, 1993.
- [POS 94] J.-G. Postaire, S. Olejnik, A relaxation scheme for improving a convexity based clustering method, *Pattern Recognition Letters*, n°15, pp. 1211-1221, 1994.
- [ROS 56] M. Rosenblatt, Remarks on some nonparametric estimates of a density function, *Ann. Math. Statist.* pp. 27, 832-837, 1956.
- [SER 82] J. Serra, Image analysis and mathematical morphology. New York : Academic, 1982.
- [TOU 89] A. Touzani, J.-G. Postaire, Clustering by mode boundary detection, *Pattern Recognition Letters*, vol. 9, pp. 1-12, 1989.
- [TOU 88] A. Touzani, J.-G. Postaire, Mode detection by relaxation, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-10, pp. 970-978, 1988.
- [VAS 80] C.P.A. Vasseur, J.-G. Postaire, A convexity testing method of cluster analysis, *IEEE Trans. Sys. Man Cybern.*, vol. SMC-10, pp. 145-149, 1980.
- [WEG 72] E.J. Wegman, Nonparametric probability density estimation, A summary of available methods, *Technometrics* 14, pp. 533-546, 1972.
- [ZHA 94] R.D. Zhang, J.-G. Postaire, Convexity dependent morphological transformations for mode detection in cluster analysis, *Pattern Recognition Letters*, vol. 27, n°1, pp. 135-148, 1994.

Manuscrit reçu le 10 novembre 1998.

L'AUTEUR

El Hassan SBAI



El Hassan Sbai a obtenu sa maîtrise d'électronique à Oujda, Maroc, en 1988 et sa thèse de 3ème cycle en automatique à Fès, Maroc, en 1992. Il a passé deux ans de stage entre 1990 et 1992 au centre d'automatique de l'université des sciences et technologies de Lille, France. Depuis 1993, il est maître assistant à l'école supérieure de technologie de Meknès, Maroc. Actuellement, il prépare son doctorat d'Etat à la faculté des sciences de Meknès. Ses intérêts scientifiques comprennent l'analyse des données, la reconnaissance des formes et la vision artificielle.