

REVUE DE STATISTIQUE APPLIQUÉE

H. EL MAY

G. ANTILLE

R. TRÉMOLIÈRES

La méthode des tranches de densité : une réponse aux problèmes de multirégression

Revue de statistique appliquée, tome 44, n° 3 (1996), p. 83-94

http://www.numdam.org/item?id=RSA_1996__44_3_83_0

© Société française de statistique, 1996, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (<http://www.sfds.asso.fr/publicat/rsa.htm>) 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/>

LA MÉTHODE DES TRANCHES DE DENSITÉ : UNE RÉPONSE AUX PROBLÈMES DE MULTIRÉGRESSION

H. El May (1), G. Antille (1), R. Trémolières (2)

(1) Université de Genève, Dpt d'économétrie

Uni Mail, Bd Carl-Vogt 102, 1211 Genève 4

(2) Université Panthéon-Assas, Paris II,

92 Rue d'Assas, 75006 Paris, France

RÉSUMÉ

Le problème que nous abordons ici est celui de la généralisation du concept de régression ayant pour but la prise en compte des aspects structurels de l'ensemble des données. Nous présentons dans cet article une méthode de résolution basée sur la percolation : la méthode des tranches de densité, une approche permettant de rechercher automatiquement, sur un même ensemble de données, l'existence d'un modèle ou de plusieurs modèles sous-jacents.

Mots-clés : régression, multirégression, régression typologique.

ABSTRACT

This paper presents a mixed method assigning a given set of points of a multidimensional space into several subsets having their own regression hyperplanes or factorial decomposition. The method is based on density concepts adapted from the Percolation Method.

Keywords : Regression, multiregression, typological regression.

1. Introduction

Dans les problèmes d'ajustement de données par des méthodes de régression, la présence de valeurs aberrantes et de certaines structures sous-jacentes peut affecter la qualité de l'ajustement. L'identification de ces particularités peut être rapprochée d'un problème de reconnaissance de formes. Dans cet article nous présentons un algorithme permettant de résoudre ce type de problèmes pour des structures linéaires.

Face à l'hétérogénéité a priori possible de l'ensemble des données nous proposons de chercher, sur un même nuage de points plusieurs ajustements et non un seul, pour permettre une description beaucoup plus précise et une meilleure compréhension du phénomène étudié.

La figure 1 montre une situation où l'évolution de y change en fonction du niveau de x . La détection du point aberrant ainsi que la détection des 4 droites, de pentes différentes, sont impératives pour une bonne description de l'ensemble des observations.

La figure 2 montre une situation encore plus complexe où le nuage des points présente deux tendances différentes qui se croisent.

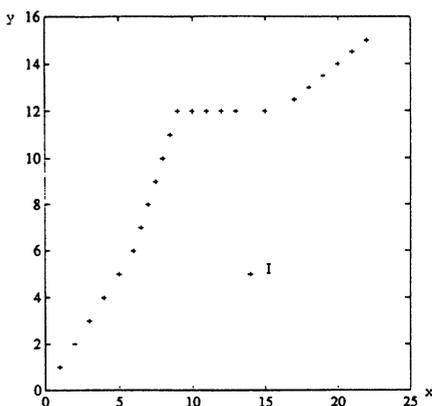


FIGURE 1

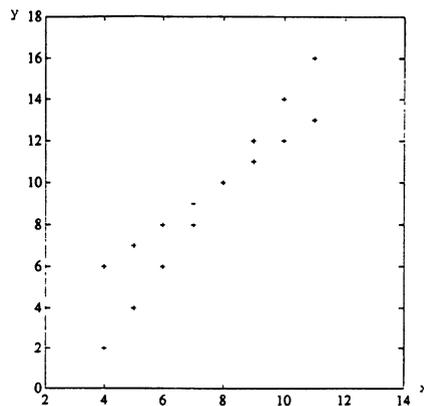


FIGURE 2

La détection de ces multitendances prend toute son importance dans un espace multidimensionnel où une représentation des points n'est plus réalisable.

Des approches multivariées, cherchant à décomposer l'ensemble des observations en groupes, traduisant chacun une relation spécifique, ont été présentées dans la littérature notamment par Mac Gee et Carleton [7], OK Sakun [6] et Charles [3].

Dans cet article nous présentons une méthode multivariée qui adapte le principe de la Percolation [9], [10], méthode de classification non hiérarchique faisant appel à la notion de densité, au cas de la régression linéaire.

Cette méthode appelée méthode des tranches de densité ([1], [2], [4], [5], [6]) permet de découvrir k tendances sous-jacentes à un même ensemble de données, k est déterminé par la méthode elle-même en fonction de la structure des observations et du degré d'affinement souhaité.

Dans la section II, une description de la méthode ainsi que l'algorithme correspondant sont présentés.

Des applications à la reconnaissance des formes, à la multirégression, ainsi qu'à la régression multiple sont présentées dans la section III; l'étude de configurations particulières y est aussi abordée.

2. La méthode des tranches de densité

Notre objectif étant de structurer un ensemble de points en fonction de leur degré d'explicativité par rapport à diverses possibilités de régression dans l'ensemble

des données, nous allons introduire diverses notions comme celle de sous-groupes de points générateurs d'un même hyperplan. Puis, pour distinguer entre les différents «hyperplans directeurs» ainsi constitués nous introduirons des notions de distances entre hyperplans dont la distinction se fera en fonction de seuils de perception comme dans la méthode de percolation. Formellement, considérons les notations et définitions suivantes afin de mieux préciser l'approche et la méthode des tranches de densité.

Soient :

$E =$ un ensemble de points de l'espace Euclidien \mathbf{R}^{p+1} , $p \geq 1$, muni d'une distance $d(\cdot, \cdot)$, où l'on affecte chaque point $i \in E$ d'un poids p_i avec $p_i \in [0; 1]$ et tel que $\sum p_i = 1$:

$$E = \{z_i = (y_i, x_i), i = 1, 2, \dots, n\} \text{ avec } x_i = (x_{i1}, \dots, x_{ip})' \text{ et } n > p.$$

$G =$ sous-ensemble arbitraire de points de E , qu'on appellera «groupe» de points pour simplifier.

$H(G) =$ hyperplan obtenu par régression sur un groupe, dans ce cas G est appelé «ensemble» ou «groupe générateur» de l'hyperplan.

On définit une tranche de mesure, – ou d'épaisseur –, η autour de $H(G)$ par

$$T(H(G), \eta) = \{z \in \mathbf{R}^{p+1} | d(z, H(G)) \leq \eta\}$$

et l'on appelle encore $H(G)$ l'hyperplan directeur de la tranche. Le sous-ensemble de points de E associés à une tranche $T(H(G), \eta)$ est noté

$$E(H(G), \eta) = E \cap T(H(G), \eta).$$

On définit la densité d'une tranche $T(H(G), \eta)$ par

$$\delta(T(H(G), \eta)) = \text{card}(E(H(G), \eta))$$

c'est-à-dire par le nombre de points de E dans la tranche $T(H(G), \eta)$. On utilisera indifféremment le terme de tendance pour parler de $H(G)$, de $T(H(G), \eta)$ ou de $E(H(G), \eta)$.

On définit la distance entre un groupe G , (non forcément groupe générateur) et un hyperplan H par

$$d(G; H) = \left(\sum_{x \in G} p_x d(x, x^*) \right) / \left(\sum_{x \in G} p_x \right)$$

où x^* , la projection de x sur H , est l'élément le plus proche de x au sens de $d(x, x^*)$ et où p_x est un poids à donner à chaque point x de G .

Pour caractériser la notion de distance entre deux groupes G_1, G_2 , ou deux tendances T_1, T_2 , au sens linéaire, on définit le concept de distance linéaire réciproque

par

$$d_{lin.rec}(G_1, G_2) = \frac{1}{2} [d(G_2, H(G_1)) + d(G_1, H(G_2))]$$

Deux groupes G_1, G_2 , ou deux tendances T_1, T_2 seront alors proches, au seuil ε et au sens de la distance linéaire réciproque si

$$d_{lin.rec}(G_1, G_2) \leq \varepsilon \quad \text{ou} \quad d_{lin.rec}(T_1 \cap E, T_2 \cap E) \leq \varepsilon .$$

On peut aussi tenir compte du nombre de points dans les groupes G_1 et G_2 , en remplaçant la moyenne arithmétique simple des 2 distances par une moyenne arithmétique pondérée par ces nombres de points. On peut aussi convenir d'autres définitions similaires.

Cette distance tient compte essentiellement de la notion d'orientation si les groupes G_1 et G_2 sont éloignés dans l'espace mais approximativement dans le prolongement l'un de l'autre, et, de la notion d'épaisseur si les groupes sont proches.

Dans la méthode des tranches de densité, nous cherchons à expliquer un caractère y relativement aux caractères $x_j, j = 1, \dots, p$ par une ou plusieurs régressions. Ainsi, chacun des points de E pourra se retrouver dans l'une ou l'autre des catégories suivantes :

- points appartenant à une même tendance (points relatifs à un même hyperplan de régression),
- «points-frontières» appartenant à plusieurs tendances simultanément,
- points isolés non rattachés à une tendance.

Dans l'application de l'algorithme, on se fixe les paramètres suivants :

- η indice d'épaisseur des tranches de départ;
- ε seuil de proximité permettant de considérer que deux tendances différentes sont suffisamment proches pour être confondues en une seule;
- g_{\min}, g_{\max} nombres ayant pour but de fixer la taille retenue des ensembles générateurs des différents hyperplans.

On note alors :

\mathcal{G} = ensemble des groupes générateurs retenus a priori dans l'algorithme.

La taille de cet ensemble peut être très grande puisqu'il peut contenir à l'extrême tous les sous-ensembles à au moins $p + 1$ éléments de E . Pour simplifier, on pourra ne s'intéresser qu'aux groupes ayant exactement $p + 1$ éléments, donc tels que $g_{\min} = g_{\max} = p + 1$.

On peut alors écrire

$$\mathcal{G} = \left\{ G^{(p+1)} \in \mathcal{P}(E) \text{ tels que } \text{card}(G^{(p+1)}) = p + 1 \right\}$$

où $\mathcal{P}(E)$ est l'ensemble des parties de E .

Lorsque la combinatoire liée à \mathcal{G} peut être considérée comme trop importante, on pourra aussi supposer que \mathcal{G} est constitué d'un sous-ensemble tiré au hasard dans l'ensemble des parties de E . Par exemple \mathcal{G} peut être supposé défini par un sous-ensemble tiré au hasard de l'ensemble des sous-ensembles à $p + 1$ éléments de E . Mais il faut bien reconnaître que la procédure aléatoire est hautement conjoncturale et qu'il convient certainement de prendre un nombre assez important de sous-ensembles pour augmenter les chances de tirer des hyperplans avec des directions proches de chaque tendance importante existante.

En effet pour qu'elle soit constituée, une sous-structure doit être représentée par au moins un groupe générateur contenu dans sa tranche et ayant une direction proche de son hyperplan directeur.

Pour des généralisations ultérieures nous introduisons en outre :

s_G = nombre minimum de points devant être spécifiques à une tendance et n'appartenant à aucune autre.

s_T = nombre minimum de points devant appartenir à la tranche associée à une tendance pour que cette tranche puisse être retenue.

Pour simplifier, on limitera l'ensemble des paramètres à des valeurs telles que

$$s = s_G = s_T = g_{\min} = g_{\max} = p + 1 .$$

Le seuil de proximité ε sera fixé à zéro au départ puis on lui fera parcourir les valeurs de la matrice de distances entre les points de l'ensemble de départ. On pourra aussi laisser le libre choix à l'utilisateur. Pour chaque valeur de ε de plus en plus élevée une nouvelle partition en $k(\varepsilon)$ groupes avec leur régime de régression associé sera donnée (pour une description plus détaillée de la méthode cf. El May, Trémolières [3], [4]). Cet ε joue l'effet de zoom et permet de déceler les structures secondaires ou principales. Les orientations internes sont de plus en plus négligées au fur et à mesure que croît le seuil de perception ε . Des mesures permettant d'apprécier la qualité des hyperplans de régression obtenus et leur pouvoir de représentation des points du groupe auquel ils sont associés sont aussi calculées.

Quant à l'indice d'épaisseur η des tranches de départ, il devrait être lié au bruit que l'on peut admettre dans la définition des groupes de départ. La distribution de probabilité des points autour de chaque hyperplan permettra de fixer ce paramètre. Cependant, par souci de simplification, nous nous contenterons de prendre $\eta = 0$, dans l'application de l'algorithme afin de limiter le temps de calcul. L'algorithme des tranches de densité peut alors être établi comme suit :

1. En fonction de la dimension $n \cdot (p + 1)$ des données, choisir \mathcal{G} comme étant l'ensemble de tous les groupes à $p + 1$ éléments ou uniquement une partie constituée de r (à fixer par l'utilisateur) groupes à $p + 1$ éléments tirés au hasard.
2. Pour tout $G \in \mathcal{G}$
 - déterminer :
 - $H(G)$, l'hyperplan de régression,
 - $T(H(G), \eta)$ la tranche à η près,
 - $\delta(T(H(G), \eta)) = \text{card}(E(H(G), \eta))$, la densité de la tranche.

- classer toutes les tranches $T(H(G), \eta)$ en ordre décroissant de leur densité. (Considérer à ce stade qu'aucune tendance n'a encore été « examinée » ou « retenue » comme constitutive de la structuration linéaire recherchée).
3. On considère alors la tendance T la plus dense non encore examinée et possédant au moins s points spécifiques, autrement dit, des points appartenant à la tendance T considérée et à aucune des tendances déjà retenues comme constitutives de la structuration linéaire recherchée.

On détermine si la tendance trouvée est proche au seuil ε , d'une tendance déjà constituée.

Si la tendance trouvée est proche d'une et une seule tendance déjà constituée, les points de la tendance sont annexés à la tendance déjà constituée.

Si la tendance se trouve voisine à un seuil inférieur ou égal à ε de deux tendances différentes, elle constitue une tendance frontière et les points spécifiques qui la constituent sont des *points-frontières*.

Dans le cas où la tendance trouvée ne possède pas de tendance voisine déjà constituée, elle est définie comme une nouvelle tendance « *constitutive* » et ses points forment un nouveau groupe.

Si l'on ne trouve pas de tendance suffisamment dense (avec au moins s points), les points en question qui ne peuvent appartenir à aucune autre tendance, sont laissés de côté comme points *isolés*.

On revient en 3. pour considérer les tendances suivantes.

4. Augmenter ε si un groupage moins fin est souhaité et retourner en 2..

Les résultats en fonction de plusieurs valeurs de ε permettent une description qui va dans le sens de la perception visuelle des données. C'est ainsi que pour les données de la figure 1, à un niveau de $\varepsilon = 0$, la méthode détecte outre le point aberrant I, les 4 tendances suivantes :

$$\begin{aligned} H_1 & : y = x \\ H_2 & : y = -6 + 2x \\ H_3 & : y = 12 \\ H_4 & : y = 4 + 0.5x \end{aligned}$$

A un niveau $\varepsilon = 2$, tous les points de la figure à l'exception du point aberrant I sont regroupés autour de la même droite: $y = 3.18 + 0.5x$. Aux niveaux intermédiaires $0 < \varepsilon < 2$, H_1 et H_2 d'un côté ainsi que H_3 et H_4 de l'autre sont regroupés.

Par ailleurs, pour l'ensemble des données décrit dans la figure 2, la méthode des tranches de densité identifie les 2 tendances suivantes au niveau $\varepsilon = 0$:

$$\begin{aligned} H_1 & : y = -6 + 2x \\ H_2 & : y = 2 + x \end{aligned}$$

Au niveau $\varepsilon = 1$, H_1 et H_2 sont regroupés autour de la droite de régression classique : $y = -2 + 1.5x$.

3. Applications

Nous avons testé la méthode des tranches de densité sur plusieurs types de configurations, sur plusieurs lettres de l'alphabet, sur des figures parfaites et sur des figures où des bruits ont été introduits autour des tendances.

Les résultats sont donnés pour différents seuils de perception ε . On initialise d'abord ε à zéro puis on lui fait parcourir la matrice des distances euclidiennes entre les points de l'ensemble de toutes les données. On peut aussi limiter l'analyse à seulement quelques niveaux de ε choisis par l'utilisateur.

Cette analyse à différents seuils permet d'avoir une masse d'information très grande qui est en même temps une source de découvertes intéressantes quant aux orientations internes et à l'emplacement des points les uns vis-à-vis des autres, mais aussi quelquefois une panoplie trop grande de résultats : l'utilisateur devra faire un arbitrage réfléchi entre, d'une part, les différents seuils de perception, le nombre de groupes, les différentes mesures de qualité des hyperplans obtenus, et, d'autre part, la structure en groupes linéaires à choisir et les résultats à retenir.

Les deux premiers exemples présentés ici sont relatifs à deux cas de reconnaissance des formes :

- l'exemple du «A parfait» où nous avons écarté tous les bruits de façon à avoir des structures parfaites (cf. Fig. 3), et,
- l'exemple du «A diffus» où nous avons introduit des bruits (cf. Fig. 4).

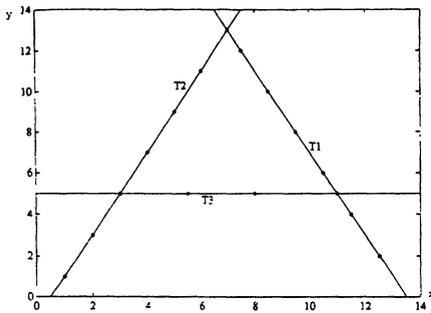


FIGURE 3

Résultats obtenus pour
 $\varepsilon = 0$, $p = 1$ et $s = 2$

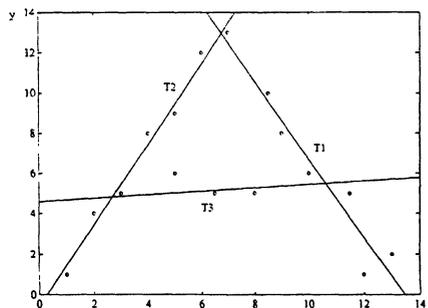


FIGURE 4

Résultats obtenus pour
 $\varepsilon = 1.118$, $p = 1$ et $s = 3$

La méthode est stable face aux bruits qui peuvent être introduits autour des tendances. Elle permet dans les deux cas de retrouver les trois barres principales du «A» : deux barres obliques et une barre horizontale.

Le troisième exemple est un nuage de points allongé où nous avons introduit un point aberrant (cf. Fig. 5) : ce point est reconnu comme tel par la méthode, il est qualifié d'isolé et écarté lors du calcul de l'hyperplan.

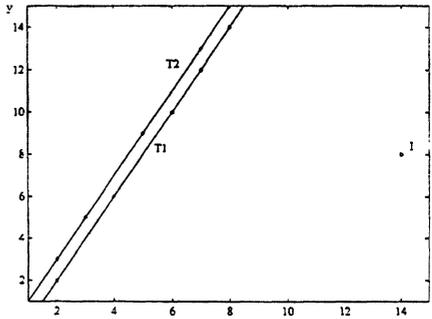


FIGURE 5a
Résultats obtenus pour
 $\varepsilon = 0$ et $s = 2$

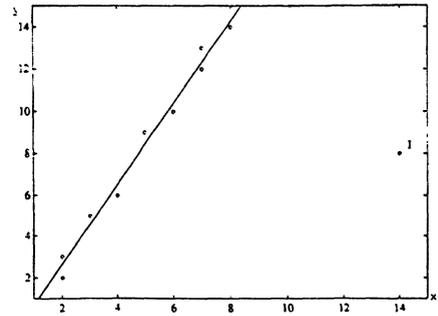


FIGURE 5b
Résultats obtenus pour
 $\varepsilon = 1$ et $s = 2$

Le quatrième exemple (cf. Fig. 6) se rapporte à un cas multidimensionnel avec 2 variables explicatives x_1 et x_2 et une variable à expliquer y . La figure est constituée de $n = 26$ points que pour la commodité nous distinguons en :

- 11 points «+» appartenant au plan $\pi_1 : y = -x_1 - x_2 + 5$,
- 6 points «o» appartenant au plan $\pi_2 : y = 2x_1 - 2x_2 + 7.5$,
- 3 points «*» appartenant au plan $\pi_3 : y = 2.5$,
- 5 points «●» appartenant aux 3 plans π_1, π_2 et π_3 , et,
- 1 point «Δ» n'appartenant ni à π_1 , ni à π_2 , ni à π_3 .

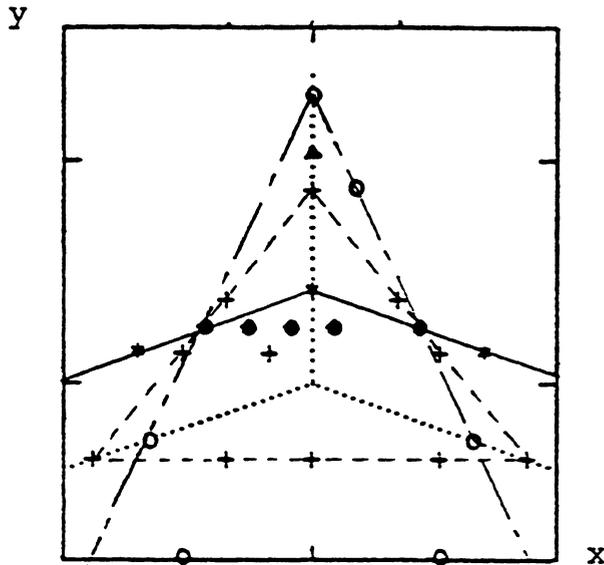


FIGURE 6
Cas multidimensionnel

Si on ne s'intéresse qu'aux groupes ayant 3 ($= p + 1$) éléments. La première itération devrait consister en une énumération de $\binom{26}{3}$ soit 2600 sous-ensembles. Pour limiter ce départ fastidieux, nous avons choisi de ne considérer que 200 sous-ensembles à 3 éléments tirés d'une manière aléatoire à partir de l'ensemble des données.

A $\varepsilon = 0$, une première tranche, la plus dense, est retenue; elle définit l'hyperplan π_1 formé des 16 points «+» et «•». Le groupe générateur de cette tranche est un des $\binom{16}{3}$ groupes composés de 3 points «+» ou «•».

Dans la constitution du second groupe générateur tous les sous-ensembles générant la première tranche ou d'autres tranches formées par moins de 3 points spécifiques sont laissés de côté par l'algorithme. On cherche alors un groupe générateur de points autres que «+» et «•» et de densité maximale. On trouve un groupe formé de trois points «o».

C'est ainsi que la deuxième tranche est constituée des 11 points «o» et «•» de π_2 .

A ce stade, il ne reste que les points «*» et le point « Δ » qui n'appartiennent pas déjà à une tranche constituée.

En écartant les points déjà inclus dans les 2 tranches précédentes, on cherche un groupe générateur de densité maximale : celui-ci est trouvé parmi les points «*». L'algorithme retient le plan π_3 formé par tous les 8 points «*» et «•».

Il reste alors un seul élément, le point « Δ » qui est considéré comme point isolé.

Ainsi au seuil $\varepsilon = 0$, l'algorithme retrouve exactement la structure existante. Pour des valeurs de ε plus grandes, $\varepsilon < 0.3$, les équations des 3 plans sont légèrement modifiées par l'affectation d'autres points de l'ensemble des données qui leur sont proches à une distance $< \varepsilon$ de façon à constituer des tranches d'épaisseur ε .

Une nouvelle tendance ressort aussi dans les résultats : c'est une tranche constituée autour d'un plan médian entre les plans π_1 et π_2 et incluant le point Δ qualifié d'isolé auparavant.

A $\varepsilon > 0.3$, le plan horizontal disparaît des résultats, ses points, excepté le point $(0, 0, 2.5)$, sont tous affectés aux autres tranches.

A $\varepsilon = 0.8$ tous les points sauf le point $(0, 0, 2.5)$ sont regroupés autour d'une seule et même tendance.

Dans une deuxième étape, la variable à expliquer y , de l'ensemble des données, a été perturbée par une erreur normale de moyenne 0 et d'écart type 0.3.

Cet écart type est arbitraire mais choisi de façon à ne pas trop compromettre la configuration de départ.

Pour des ε petits inférieurs ou égaux à 0.4, les points sont regroupés par la méthode autour de plans très proches de π_1 , π_2 et du plan médian. Le plan $y = 2.5$ ne ressort plus dans les résultats puisque tous ses points à l'exception d'un seul sont affectés aux autres plans.

Pour des valeurs de ε plus grandes, la méthode des tranches de densité forme un groupe et détecte un point isolé, le point $(0, 0, 2.5)$. Enfin la méthode des tranches de densité regroupe tous les points dans une même tranche autour du plan de régression des moindres carrés : $y = 5.52 - 1.2 x_1 - 1.2 x_2$.

La méthode permet une bonne description de l'ensemble des données surtout si l'on tient compte des différents résultats liés à chaque niveau de ε .

Dans les exemples suivants, nous présentons le comportement de la méthode face à certaines configurations particulières.

Pour la représentation illustrée à la figure 7, la méthode identifie des tendances horizontales et non verticales puisqu'elle tient compte non seulement de la notion d'orientation mais aussi de la densité en nombre de points qui est utilisée pour choisir les candidats aux groupes.

Pour l'illustration 8, plusieurs sous-groupes de points de plusieurs tranches sont regroupés par la méthode autour de la même tendance. La tendance T_1 est formée en premier parce qu'elle est plus dense que T_2 et T_3 . Par contre, si cette tendance était formée au total par un certain nombre de points plus petit que le nombre de points de T_2 et celui de T_3 , ces deux tendances seraient formées en premier, et les différentes masses n'auraient pas pu former un nouveau groupe par manque de points propres, les points en question étant déjà classés dans des groupes.

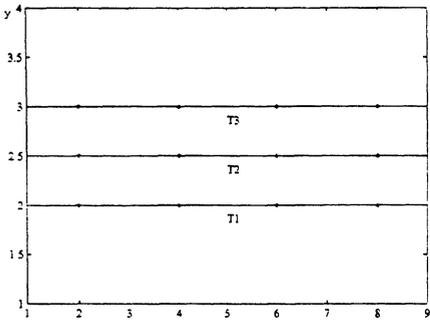


FIGURE 7

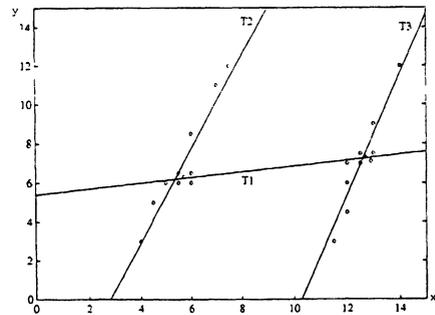


FIGURE 8

4. Conclusion

Les domaines d'applications de la méthode des tranches de densité sont nombreux en reconnaissance de formes à structures linéaires ou en régression multiple et multirégression linéaires.

La méthode nous paraît fiable face aux représentations où les tendances sont assez nettes et non ambiguës. A différents seuils de perception la méthode décrit le nuage de points initial, en définissant des tendances « naturelles » linéaires en harmonie avec la perception visuelle.

Cependant, les groupes formés n'obéissent pas forcément au concept de continuum, un concept non pris en compte par la méthode. Plusieurs sous-ensembles

de points appartenant à différentes tranches peuvent être regroupés par la méthode autour de la même tendance s'ils ont plus ou moins la même orientation. Ainsi on risque de mettre dans la même tranche des groupes de points assez distants mais formant le même hyperplan de régression. Les tranches obtenues peuvent être repérées en calculant un indice mesurant la cohésion et l'homogénéité des groupes trouvés. On pourrait se faire une idée de la distribution des points et des éventuelles concentrations de points dans les tranches, en projetant par exemple les points de la tranche sur l'hyperplan directeur de celle-ci et en représentant si possible la distribution des points obtenue ou en calculant un indice. C'est à l'utilisateur par la suite de décider, en fonction du problème traité, s'il est opportun de tenir compte des groupes différents qui peuvent être ainsi obtenus.

Les problèmes à traiter pouvant être très disparates, il faudrait essayer de donner une plus large place à l'heuristique et à l'interactif, en utilisant le bon sens de l'utilisateur qui devra intervenir même au niveau de la résolution du problème afin de guider le déroulement de l'algorithme. Ainsi on pourrait prévoir d'intégrer différentes définitions de proximité des tranches qui seraient choisies en fonction du problème traité, en privilégiant tantôt la notion d'orientation, tantôt la notion de cohésion, tantôt la notion d'épaisseur.

Dans le cas où le concept de continuum peut être privilégié, on peut se contenter de déterminer des «groupes» homogènes, par exemple au sens de la percolation et définir des tendances propres à chacun d'eux, pourvu évidemment que ces groupes forment des tendances suffisamment denses.

Dans le cas où le concept de tendance doit être privilégié, la méthode des tranches de densité permet par contre d'aborder le problème de façon plus générale en offrant la possibilité d'obtenir des hyperplans de régression constitués de groupes différents mais relatifs à des tendances identiques.

Références

- [1] Antille G., El May H. (1992). The Use of Slices in the LMS and the Method of Density Slices : Foundation and Comparison, in Dodge Whittaker (Ed.), *Computational Statistics*, Vol. 1 (Physica-Verlag) 441-445.
- [2] Antille G., El May H. (1994). Méthodes Descriptives en Régression Linéaire, *XXVI^e Journées de Statistique de l'ASU*, Neuchâtel 1994.
- [3] Charles C. (1977). La Régression Typologique et la Reconnaissance des Formes, *Thèse de 3^{ème} cycle*, Paris IX.
- [4] El May H. (1991). Extension de la Régression Classique à des Problèmes Typologiques et Présentation de la Méthode des Tranches de Densité : une Approche Basée sur la Percolation, *thèse de doctorat*, IAE. Université d'Aix-Marseille III.
- [5] El May H. (1992). La Méthode des Tranches de Densité : une Réponse aux Problèmes de Multirégression, *cahier 92.05 du département d'économétrie de l'Université de Genève*.

- [6] El May H. and Trémolières R. (1992). Presentation of the Method of Density Slices : An Answer to the Multitendencies Problems, in Dodge Whittaker (Ed.), *Computational Statistics*, Vol. 1 (Physica-Verlag) pp. 171-178.
- [7] OK Sakun Y. (1975). *Analyse Factorielle Typologique et Lissage Typologique. thèse de 3e cycle, Université de Paris VI.*
- [8] Mac Gee U.E., Carleton W.T. (1970). Piecewise Regression. *JASA*, September 70, Vol. 65, No 331, Applied Section, pp. 1109-1124.
- [9] Trémolières R., (1983) The Generalized Percolation for Data Analysis and Pattern Recognition. In «New Trends in Data Analysis and Applications», J. Jansen, J.F. Marcotorchino, J.M. Proth, ed., *North Holland*, New York, pp. 125-145.
- [10] Trémolières R. (1994). Percolation and Multimodel Data Structuring. In «New Approaches in Classification and Data Analysis» (Diday and al. eds), *Springer-Verlag*.