Institutional Repository - Research Portal
Dépôt Institutionnel - Portail de la Recherche
researchportal.unamur.be
University of Namur

THESIS / THÈSE

MASTER EN SCIENCES MATHÉMATIQUES

Une nouvelle méthode de classification basée sur le processus de Poisson non homogène

Ochao, Saïda

Award date:
1992

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 03. Jul. 2025

FACULTES UNIVERSITAIRES NOTRE-DAME DE LA PAIX

NAMUR

FACULTES DES SCIENCES

André HARDY
Département de Mathématique
Facultés Universitaires de Namur
Rempart de la Vierge 8
B - 5000 NAMUR - BELGIUM

UNE NOUVELLE METHODE DE CLASSIFICATION BASEE SUR LE PROCESSUS DE POISSON NON-HOMOGENE

Promoteur: J-P. RASSON

Saïda OCHAO

Année académique 1991 - 1992

REMERCIEMENTS

Avant de confier mon travail au lecteur, je tiens tout d'abord à remercier Monsieur RASSON de m'avoir accueillie comme mémorisante, et de m'avoir fourni son aide, ses conseils et sa disponibilité tout au long de cette année.

Enfin, j'aurai une pensée particulière pour mes parents, ainsi que pour toutes les personnes qui, par leur présence, m'ont encouragée et soutenue dans mes efforts.

P L A N G L O B A L
****

Introduction

Formalisation du problème

Première partie: Processus de Poisson homogène

1. Introduction
2. Propriétés fondamentales du Processus de Poisson homogène
3. Estimation du maximum de vraisemblance d'un convexe D sur base d'une réalisation d'un Processus de Poisson homogène sur D
4. Analyse discriminante - Premier critère: domaines disjoints
5. Analyse discriminante - Second critère: domaines non-disjoints
6. Conclusion

Deuxième partie: Processus de Poisson non homogène

1. Introduction
2. Estimateur du maximum de vraisemblance d'un convexe D sur base d'une réalisation d'un Processus de Poisson non homogène sur D
3. Analyse discriminante - Premier critère: domaines disjoints
4. Analyse discriminante - Second critère: domaines non-disjoints
5. Conclusion

Troisième partie: Méthodes d'estimation de densité

1. Introduction
2. Methodes non paramétriques
3. Fonction noyau K(x)
4. Noyau de l'estimateur naïf appliqué au cas discret

Quatrième partie: Evaluation de la nouvelle méthode de classification pour le traitement d'images en télédétection

1. Introduction
2. Principe du traitement d'images par classification
3. Problématique du traitement d'images par classification
4. Exemples

Conclusion

Références et Bibliographies

Table des matières

INTRODUCTION

L'objet de ce mémoire constitue une nouvelle réflexion dans le domaine de la classification supervisée.

Cette recherche est la continuation de travaux entrepris par Pierre BAUFAYS (1985); où celui-ci proposait une règle d'affectation de points suivant l'hypothèse d'un Processus de Poisson homogène. Il supposait que les points à classifier étaient tels que leurs signatures spectrales étaient distribuées uniformément et indépendamment dans des domaines convexes et disjoints.

Le critère qu'il découvrait en minimisant le Risque BAYESIEN de mauvaise affectation était d'affecter le point à classer à l'échantillon étiqueté auquel il ajoutait, par convexité, un hypervolume minimal.

L'intérêt particulier du critère était l'introduction par HARDY et RASSON (1982) de l'utilisation de la mesure de LEBESGUE comme mesure de proximité alors que toutes les méthodes connues utilisaient des distances ou semi-distances.

Mais, malheureusement le modèle théorique ne permettait pas de classifier les points appartenant à l'intersection des enveloppes convexes de plusieurs bases d'entrainement; ceux-ci étaient classifiés selon une règle empirique.

Afin d'obtenir un modèle théorique pour le schéma entier d'une classification, nous allons généraliser l'hypothèse du Processus de Poisson homogène en un Processus de Poisson non homogène.

L'hypothèse d'une distribution uniforme est abandonnée, laissant la place à: une adéquation des données à n'importe quelle densité non paramétrique.

Nous gardons le fait que les points des bases d'entrainement sont répartis dans des domaines convexes, mais que nous n'avons plus besoin de supposer nécessairement disjoints.

La minimisation du risque bayesien nous donne, alors, une seule règle théorique pour tous les points à classer.

Nous consacrerons une partie du mémoire à l'estimation de densité.

En effet, notre modèle théorique se base sur des intensités inconnues qu'il faudra estimer. En pratique, nous utiliserons comme outil pour estimer ces intensités, le noyau de l'estimateur naïf de densité proposé par B.W. SILVERMAN.

Pour terminer ce travail, nous évaluerons notre nouvelle méthode de classification aux traitements d'images en télédétection.

...

FORMALISATION DU PROBLEME

Un expérimentateur observe un individu qu'il sait appartenir à une parmi K catégories et il désire déterminer celle dont cet individu est issu.

Pour classer cet individu, nous disposons des informations suivantes:

- des valeurs de n mesures effectuées sur cet individu. On suppose que ces valeurs sont quantitatives.
  On peut donc associer à cet individu, un vecteur x d'un sous-espace V de IR^n

  $$\text{individu} \longrightarrow x = (x_1, x_2, \ldots, x_n) \text{ de } V \subset IR^n$$

- de K ensembles d'individus utilisés comme éléments de référence. Il y a un ensemble par catégorie. Chaque ensemble i est composé de n_i individus (i= 1 ... K).
  Pour chaque individu des différents groupes de référence, on dispose des mêmes variables que pour l'individu à classer. L'ensemble E de tous les individus de référence est appelé "ensemble de référence".

  $$E = (x_{ij}, j = 1 \ldots n_i, i = 1 \ldots K)$$

  On peut associer, à chaque élément x_{ij} de E, un vecteur d'un sous-espace V de IR^n

Il faut remarquer que l'appartenance des individus de E à leurs groupes respectifs n'est jamais remise en question en analyse discriminante.

Ce que nous désirons, c'est pouvoir associer à tout élément de V, une classe parmi les K catégories de référence.

Nous allons donc définir l'application d, appelée "règle de décision" telle que:

$$d: V \longrightarrow (1, 2, \ldots, K)$$
$$x \longrightarrow d(x)$$

...

De manière générale, elle dépend des individus étiquetés x_{ij} $(j = 1 \ldots n_i, i = 1 \ldots K)$, et elle est définie de manière à commettre le moins d'erreurs de classement possible. La règle d détermine une Partition $(P_1, \ldots P_K)$ de V

$$P_i = (x \text{ t q } d(x) = i) \quad i = 1 \ldots K$$

c'est-à-dire,

P_i est l'ensemble des points qui seront affectés par la règle de décision d à la catégorie i de référence.

La règle peut être définie au moyen de K fonctions $h_i(x)$, $i = 1 \ldots K$

La fonction $h_i(x)$, associée au groupe i, mesure la "ressemblance" entre x et les individus de $X_i = (x_{ij} \quad j = 1 \ldots n_i)$

Dès lors:

$$P_i = (x: h_i(x) \geqslant h_j(x) \; j = 1 \ldots K) \quad i = 1 \ldots K$$

L'ensemble $D_{ij} = (x: h_i(x) = h_j(x))$ est appelé surface de décision.

D_{ij} constitue la frontière entre P_i et P_j, lorsqu'on ne considère que ces deux populations. En présence de K catégories, il existe $K(K-1) \setminus 2$ surfaces de décisions.

...

PREMIERE PARTIE: PROCESSUS DE POISSON HOMOGENE

1. Introduction

Avant d'entamer le problème de classification concernant le Processus de Poisson non homogène, nous allons exposer les différentes étapes de la démarche poursuivie dans l'analyse discriminante suivant un Processus de Poisson homogène.

Les hypothèses de la méthode proposée par RASSON supposent que la population à classifier est une réalisation d'un Processus de Poisson stationnaire; dans l'union de K domaines convexes disjoints D_i ; dès lors à chaque D_i correspondra une classe de la population P_i

Dans cette situation, le critère de classement dépend de la mesure naturelle de l'espace, à savoir:

la mesure de LEBESGUE

Nous allons nous remémorer le raisonnement.

...

2. Propriétés fondamentales du Processus de Poisson homogène

2.1. Hypothèses

Un échantillon $\bar{X}$ est une réalisation d'un Processus de Poisson homogène sur un domaine D, si les hypothèses suivantes sont vérifiées:

- la variable aléatoire N(D) qui compte le nombre de points dans le domaine D suit une distribution de Poisson, dont le paramètre est la mesure de LEBESGUE m(D) de D; c'est-à-dire nous supposons que le nombre moyen de points dans D est proportionnel à m(D).

- les variables aléatoires N(D) et N(D') sont indépendantes si D et D' sont disjoints.

2.2. Propriété

Il découle des hypothèses du Processus de Poisson homogène, la propriété suivante:

- Conditionnellement au fait que n points aléatoires engendrés par ce processus appartiennent au domaine D, ceux-ci sont distribués uniformément et indépendamment dans D.

2.3. Remarque

Dans la suite de cette première partie, tout individu de l'échantillon $\bar{X}$ suivra une distribution uniforme dans D; ceci implique que la fonction de densité s'écrit:

$$f_D(x) = \frac{1}{m(D)} 1_D(x) \quad \forall x$$

où $1_D(x)$ est la fonction indicatrice du domaine D, à savoir:

$$1_D(x) = 1 \quad x \in D$$
$$= 0 \quad x \notin D$$

...

. m(D) est la mesure de LEBESGUE de D qui dépend de la dimension de l'espace de travail, à savoir:

- sur la droite réelle: c'est la distance usuelle;
- dans le plan: c'est la surface;
- dans l'espace à trois dimensions: c'est le volume;
- dans un espace de dimension supérieure: c'est l'hypervolume.

...

3. Estimateur du maximum de vraisemblance d'un convexe D sur base d'une réalisation d'un Processus de Poisson homogène sur D

3.1. Vraisemblance du modèle

Soit une réalisation $\overline{X} = (X_1, \ldots, X_n)$ d'un tel processus. Celle-ci sera constituée de n points indépendants et identiquement distribués, de densité $f_D(x)$ égale à:

$$f_D(x) = \frac{1}{m(D)} 1_D(x) \quad \forall x$$

Dès lors, nous pourrons écrire la vraisemblance du vecteur $\overline{X}$ sur D:

$$f_D(\overline{X}) = \prod_{i=1}^{n} f_D(X_i)$$

$$= \prod_{i=1}^{n} \frac{1}{m(D)} 1_D(X_i)$$

$$= \frac{1}{[m(D)]^n} \prod_{i=1}^{n} 1_D(X_i)$$

Par hypothèse, le domaine D est un convexe; ceci implique que si nous considérons le plus petit convexe contenant les n points, c'est-à-dire l'enveloppe convexe $H(\overline{X})$ de notre échantillon, la vraisemblance aura comme expression:

$$f_D(\overline{X}) = \frac{1}{[m(D)]^n} 1_D(H(\overline{X})) \qquad (*)$$

...

3.2. Estimation du convexe D

Nous pouvons tirer deux constatations de l'expression (*):

1) l'estimateur doit au moins contenir $H(\overline{X})$ pour que la fonction indicatrice du domaine soit non nulle;

2) l'estimateur ne peut pas être plus grand que $H(\overline{X})$ sinon la mesure de LEBESGUE correspondante serait également supérieure à celle de $H(\overline{X})$; ceci implique que nous ajoutons une quantité positive au dénominateur qui deviendrait plus grand et par conséquent, la vraisemblance serait plus petite.

En conclusion, l'estimateur du maximum de vraisemblance du domaine convexe D sur base d'une réalisation d'un Processus de Poisson homogène sur D est l'enveloppe convexe $H(\overline{X})$ de cette réalisation.

...

4. Analyse discriminante - PREMIER CRITERE: domaines disjoints

4.1. Fonction de densité en un point du domaine

Soit $\overline{X} = (X_1, \ldots, X_n)$ la réalisation d'un Processus de Poisson homogène sur un domaine convexe D tel que

$$D = \bigcup_{j=1}^{K} D_j$$

où $(D_j)_{j = 1 \ldots K}$ sont des domaines convexes et disjoints.

Tout individu de l'échantillon suit une distribution dans D dont la densité $f_D(x)$ est égale à:

$$f_D(x) = \frac{1}{m(D)} 1_D(x) \quad \forall x$$

Comme D est l'union de K domaines disjoints D_j $(j = 1 \ldots K)$, la densité $f_D(x)$ peut se ré-écrire sous la forme:

$$f_D(x) = \frac{\sum_{j=1}^{K} 1_{D_j}(x)}{m\left(\sum_{j=1}^{K} D_j\right)}$$

$$= \frac{\sum_{j=1}^{K} 1_{D_j}(x)}{\sum_{j=1}^{K} m(D_j)}$$

Posant:

. p_i : la probabilité à priori qu'un individu appartienne au domaine D_i, à savoir: $p_i = \frac{m(D_i)}{m(D)}$

...

. $f_i(x)$: la densité conditionnelle de x sachant que x appartienne au domaine D_i, à savoir:

$$f_i(x) = \frac{1}{m(D_i)} 1_{D_i}(x)$$

Nous obtenons que la densité $f_D(x)$ a pour expression:

$$f_D(x) = \sum_{i=1}^{K} p_i f_i(x)$$

Remarques:

Il est à noter que:

1) l'expression de $f_i(x)$ et de p_i suppose que: tout individu de la $i^{ème}$ population est uniformément distribué dans le domaine convexe D_i et que la probabilité à priori, qu'il appartienne à cette catégorie, est proportionnelle à la mesure de LEBESGUE de D_i.

2) le paramètre inconnu de la $i^{ème}$ population est le domaine convexe D_i lui-même. Rappelons, pour notre propos, que l'estimateur du maximum de vraisemblance de D_i est l'enveloppe convexe des individus de ce groupe.

3) la règle d'affectation Bayesienne ne pourra s'appliquer que lorsque les enveloppes convexes $H(X_1 \ldots X_K)$ sont disjointes car $(D_i)_{i=1 \ldots K}$ sont des convexes disjoints.

...

4.2. Règle de décision

4.2.1. Risque associé à chaque population

La décision d'affecter un individu à une certaine classe peut comporter des erreurs de classement, ainsi nous pouvons affecter erronément un individu du groupe i(i = 1 ... K) dans une classe j (j = 1 ... K, j $\neq$ i).

La probabilité de commettre une telle erreur est égale à:

$$P_{ji} = \int_{P_j} f_i(x)dx;\ j \neq i, j = 1 \ldots K,\ i = 1 \ldots K$$

où $f_i(x)$: densité conditionnelle de x sachant que x appartient à D_i

. $P_j = (x: d(x) = j)$

A chaque population i est associé le risque $R_i(d)$ de la mauvaise affectation d'un individu de celle-ci, par la règle de décision d:

$$R_i(d) = \sum_{j \neq i} P_{ji} \qquad j = 1 \ldots K, i = 1 \ldots K$$

La règle de décision optimale serait celle pour laquelle tous les risques sont nuls. Mais cette situation n'est pas courante en pratique.

4.2.2. Risque bayesien

Supposons que chacune des probabilités à priori, qu'un individu appartienne à un groupe donné soit connue (p_i, i = 1 ... K). Nous choisirons la règle qui minimise le risque bayesien R(d):

$$R(d) = \sum_{i=1}^{K} p_i R_i(d)$$

...

Développons cette expression.

$$R(d) = \sum_{i=1}^{K} p_i \sum_{\substack{j=1 \\ j \neq i}}^{K} \int_{P_j} f_i(x)dx$$

$$= \sum_{i=1}^{K} p_i \left[\left(\sum_{j=1}^{K} \int_{P_j} f_i(x)dx \right) - \int_{P_i} f_i(x)dx \right]$$

$$= \sum_{j=1}^{K} \int_{P_j} \sum_{i=1}^{K} p_i f_i(x) \; dx - \sum_{i=1}^{K} \int_{P_i} p_i f_i(x)dx$$

$$= \sum_{j=1}^{K} \int_{P_j} \sum_{i=1}^{K} p_i f_i(x)dx - \sum_{j=1}^{K} \int_{P_j} p_j f_j(x)dx$$

$$= \sum_{j=1}^{K} \int_{P_j} \left[\sum_{i=1}^{K} p_i f_i(x) - p_j f_j(x) \right] dx$$

$$= \sum_{j=1}^{K} \int_{P_j} (f(x) - p_j f_j(x)) \, dx$$

Pour chaque j, cette quantité ne peut être plus petite que si

$$P_j = (x \text{ t q } p_j f_j(x) \geqslant p_k f_k(x) \quad \forall k \neq j, k = 1 \ldots K) \; j = 1 \ldots K$$

...

Remarques:

1) Notons que les fonctions discriminantes $(h_j)_{j\ j=1\ldots K}$ de la règle de décision d

sont définies par:

$$h_j(x) = -\left[\sum_{i=1}^{K} p_i f_i(x) - p_j f_j(x)\right]$$

Nous obtenons bien que pour chaque j

$$P_j = (x \text{ t.q } h_j(x) \geqslant h_k(x);\ \forall k \neq j,\ k = 1 \ldots K)\ j = 1 \ldots K$$

$$\Updownarrow$$

$$P_j = (x \text{ t.q } p_j f_j(x) \geqslant p_k f_k(x),\ \forall k \neq j,\ k = 1 \ldots K)\ j = 1 \ldots K$$

2) L'ensemble des individus pour lesquels il existe au moins deux indices distincts j et k pour lesquels

$$p_j f_j(x) = p_k f_k(x)$$

est de mesure nulle, lorsque les densités f_i sont absolument continus. L'affectation de ces individus ne perturbe pas le risque bayesien; celle-ci peut donc satisfaire une règle conventionnelle choisie par l'utilisateur, par exemple:

$$d(x) = \min\ (j : p_j f_j(x) \geqslant p_k f_k(x)\ \ \forall k \neq j\ \ k = 1 \ldots K)$$

4.2.3. Conclusion

La loi de décision correspondant à notre modèle sera donnée par la règle d'affectation Bayesienne où les paramètres inconnus,

$(D_i)_{i\ i=1\ldots K}$, seront remplacés par leur

estimateur de maximum de vraisemblance.

Rappelons que le domaine d'affectation à la population i est donné par

$$(x: p_i f_i(x) \geqslant p_k f_k(x)\ \ \forall k \neq i\ \ k = 1 \ldots K)\ i = 1 \ldots K$$

...

4.2.3.1. Premier cas: le point x_o, à classer, est à l'extérieur des enveloppes convexes $H(X_i)$ et $H(X_k)$

. Si le point x_o est classé dans la $i^{ème}$ catégorie, nous estimons les domaines D_j de la façon suivante:

$$\hat{D}_j = H(X_i \vee (x_o)) \quad j = i;$$

$$= H(X_j) \quad j \neq i;$$

et

$$\hat{D}_j(x_o) = 1 \quad \text{si } j = i;$$

$$= o \quad \text{si } j \neq i;$$

car x_o ne peut appartenir qu'à un seul domaine.

. Dès lors, l'estimateur du maximum de vraisemblance de $p_i f_i(x_o)$ est donné par:

$$\hat{p_i f_i}(x_o) = \left[i \Big/ \sum_{j=1}^{K} m(H(X_j)) + S_i(x_o) \right]$$

où

$$S_i(x_o) = m(H(X_i \vee (x_o))) - m(H(x_i))$$

D'autre part, nous avons supposé les domaines D_j disjoints, nous ne pouvons donc pas allouer x_o au $i^{ème}$ groupe, si $H(X_i \vee (x_o))$ possède une intersection non vide avec au moins un ensemble $H(X_j)$ $\forall j \neq i$

Si c'est le cas, nous poserons $S_i(x_o) = +\infty$

...

. La règle bayesienne (1) nous permet d'écrire
"x_o est affecté à la population i"

ssi

$$\hat{p}_i \hat{f}_i(x_o) \geqslant \hat{p}_k \hat{f}_k(x_o) \quad \forall k \neq i$$

Ceci nous donne:

$$\sum_{j=1}^{K} m(H(X_j)) + S_i(x_o) \leqslant \sum_{j=1}^{K} m(H(X_j)) + S_k(x_o) \quad \forall k \neq i; (2)$$

$$\text{et } H(X_i \cup (x_o)) \cap H(X_j) = \emptyset \quad \forall j \neq i;$$

où dans ce cas

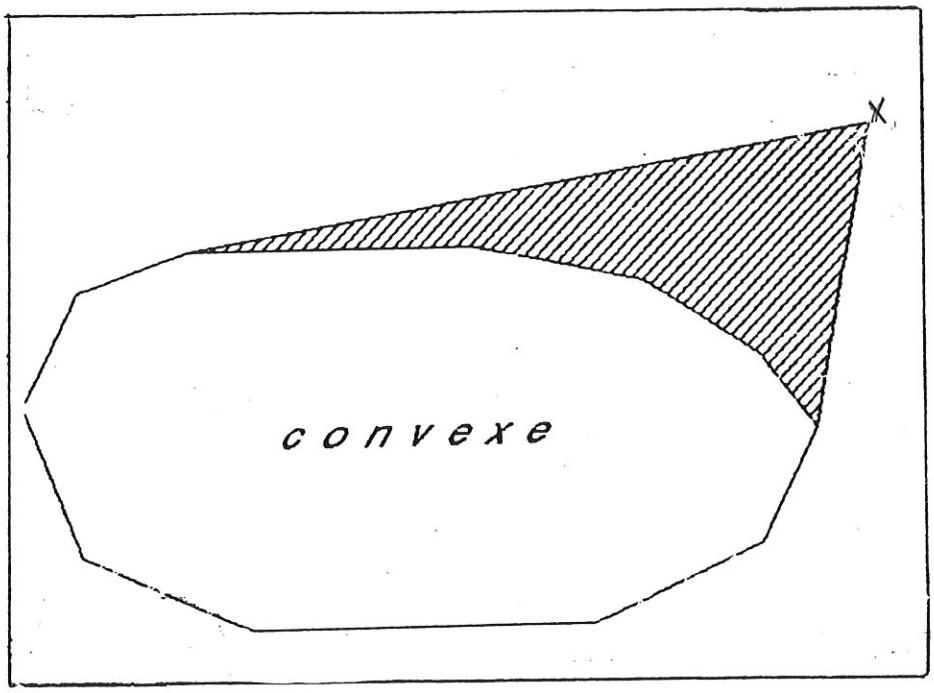
$$S_i(x_o) = m(H(X_i \cup (x_o))) - m(H(X_i)),$$

c'est-à-dire,

$S_i(x_o)$ représente la mesure de LEBESGUE que x_o "ajoute" à l'enveloppe convexe de l'échantillon étiquetté dû groupe i, lorsqu'il y est incorporé.

...

La figure suivante montre ce que représente l'hypervolume "ajouté" à une enveloppe convexe.



Conclusion:

En posant

$$S_i(x_o) = m(H(X_i \cup (x_o))) - m(H(X_i)) \quad \text{si } H(X_i \cup (x_o)) \cap H(X_j) = \emptyset$$

$$\forall j \neq i;$$

$$= +\infty \qquad \text{sinon;}$$

Nous tirons de la relation (2), la règle de classement suivante:

"le domaine d'affectation à la population i" est tel que:

$$P_i = (x : S_i(x_o) \leqslant S_k(x_o) \quad k = 1 \ldots K) \quad i = 1 \ldots K$$

Par conséquent, si le point x_o, à classer, se trouve à l'extérieur des enveloppes convexes, le point est affecté à la classe pour laquelle la mesure de LEBESGUE ajoutée par convexité est minimale.

...

4.2.3.2. Second cas: le point x_o , à classer, appartient à une seule enveloppe convexe.

Du cas précédent, il découle qu'il faut classer x_o dans le groupe correspondant puisque l'on n'ajoute plus de mesure. Ainsi, nous rencontrons la première condition d'admissibilité de FISHER et VAN NESS.

"Si les enveloppes convexes $H(X_i)$ sont disjointes et x appartient à une de ces enveloppes convexes, alors x devrait être affecté à la classe correspondante".

...

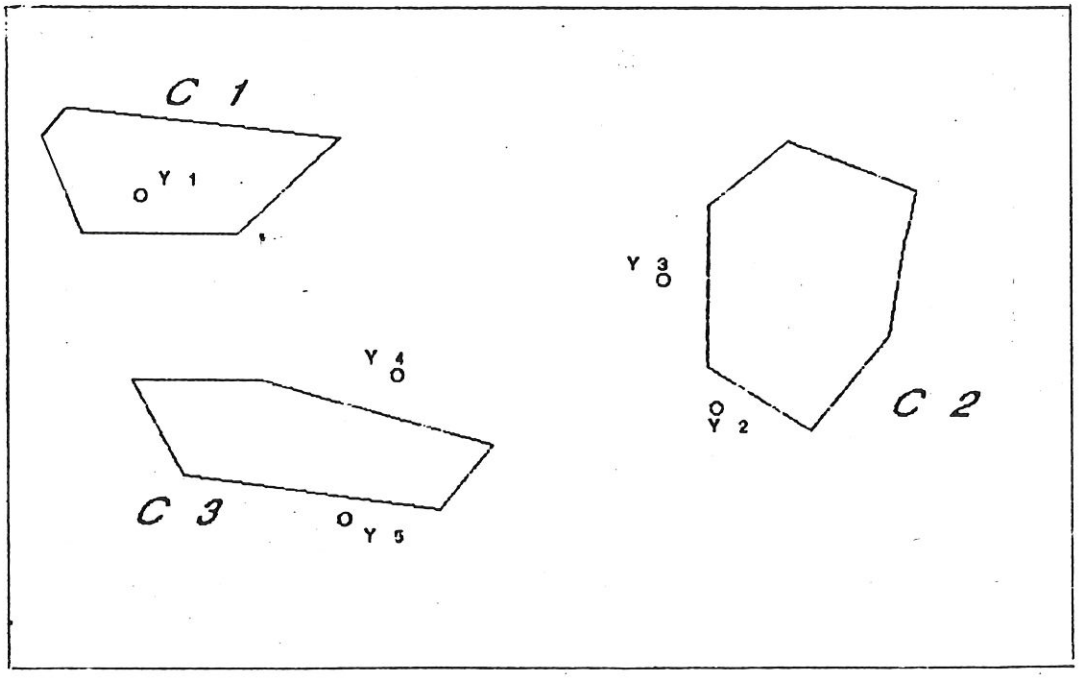
4.2.4. Exemple

Nous disposons de trois classes de référence définissant trois convexes

C_i (i = 1 ... 3)

Le problème consiste à classer les points y_j (j = 1 ... 5)

parmi les trois classes.
Si nous appliquons la règle de décision, définie plus haut, l'individu y_1 sera affecté au groupe 1 car il appartient à l'enveloppe convexe de l'échantillon étiquetté de ce groupe.
Les individus y_2 et y_3 (respectivement y_4 et y_5) seront affectés à la catégorie 2 (respectivement 3) en fonction des mesures ajoutées.



...

5. Analyse discriminante - SECOND CRITERE: domaines non disjoints

Chaque individu du domaine D_i possède:

1) une fonction de densité conditionnelle $f_i(x)$ qui s'exprime par:

$$f_i(x) = \frac{1}{m(D_i)}\, D_i(x)$$

2) une probabilité à priori que x appartienne à ce domaine:

$$p_i = \frac{m(D_i)}{m(D)}$$

5.1. PREMIER CAS: le point x_o n'appartient à aucune enveloppe convexe. Dans ce cas, nous appliquerons la règle d'affectation bayesienne:

x_o est affecté à la classe j pour laquelle

$$S_j(x_o) = m(H(X_j \cup (x_o))) - m(H(X_j))$$

est minimal.

5.2. SECOND CAS: le point x_o appartient à une seule enveloppe convexe $H(X_i)$. D'après la règle d'affectation bayesienne, nous affecterons x_o à la classe correspondante, puisque

$$\hat{p}_j f_j(x_o) = 1 \Big/ \sum_{k=1}^{K} m(H(X_k)) \qquad j = i$$

$$= 0 \qquad \text{sinon}$$

...

5.3. TROISIEME CAS: le point x_o appartient à k ($1 \leqslant k \leqslant K$) enveloppes convexes:

$$H(X_{i_1}), \ldots, H(X_{i_k})$$

Nous obtenons:

$$\hat{p}\, f_{j\, j\, o}(x) = 1 \Big/ \sum_{i=1}^{K} m(H(X_i)) \qquad j = i_1, \ldots, i_k$$

$$= 0 \qquad \text{sinon}$$

Ici, il nous est impossible de classer x_o dans l'un des domaines convexes $D_{i1}, \ldots, D_{ik}$

Exemple:

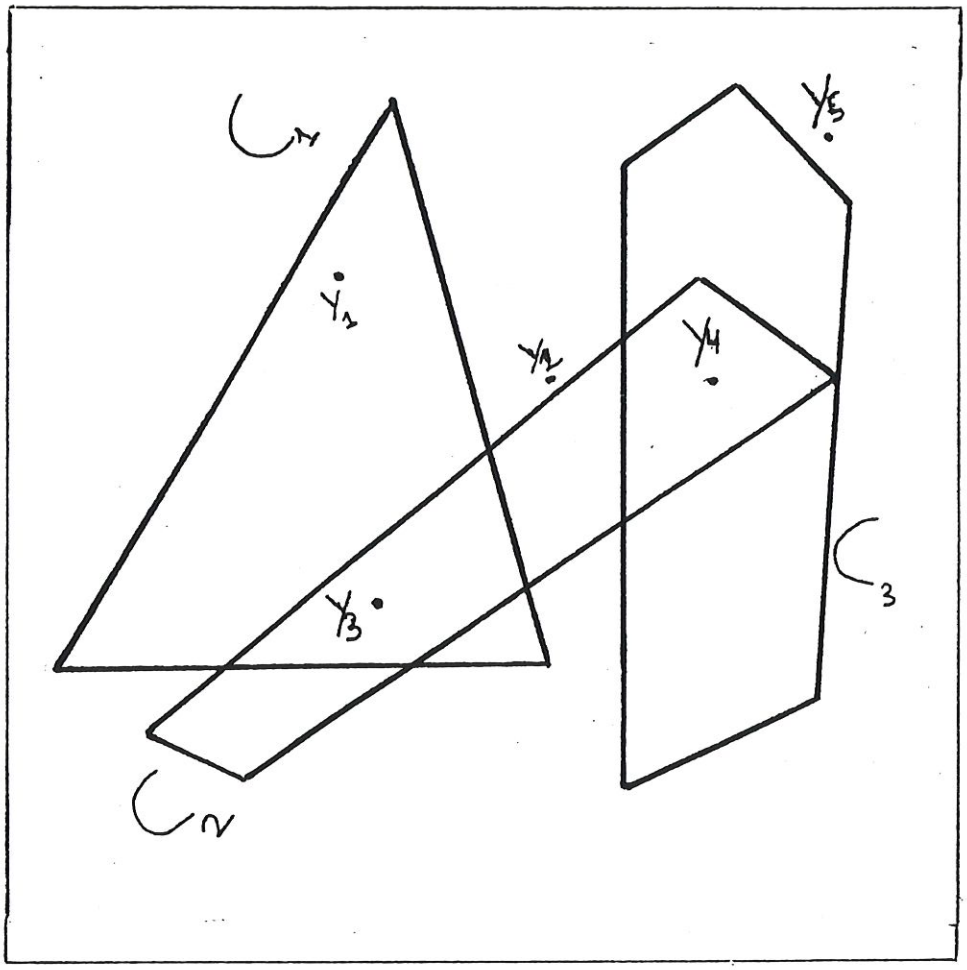
Regardons la figure suivante de la page 24:

y_1 est affecté au groupe 1 du fait de son appartenance à une seule enveloppe convexe,

y_2 est classé dans la seconde catégorie ainsi que y_5 est classé dans le troisième groupe sur base de "la mesure de LEBESGUE ajoutée"

y_3 (respectivement y_4) est un cas d'indécision entre les populations 1 et 2 (respectivement 2 et 3).

...


C1
Y1
Y2
Y3
Y4
Y5
C3
C2

...

6. Conclusion

Notre démarche, basée sur le Processus de Poisson homogène, ne nous permet pas de classer des points appartenant à l'intersection des enveloppes convexes de plusieurs bases d'entrainement. Il nous faut, dès lors, envisager une nouvelle hypothèse, à savoir:

le Processus de Poisson non homogène.

...

DEUXIEME PARTIE: PROCESSUS DE POISSON NON HOMOGENE

1. Introduction

Nous allons généraliser l'hypothèse du Processus de Poisson homogène en un Processus de Poisson non homogène.

Cette condition implique que les distributions conditionnelles au fait que n points aléatoires engendrés par ce processus appartiennent au domaine D, sont absolument quelconques.

Le critère de classement ne dépend plus de la mesure de LEBESGUE, mais de l'intensité du processus.

...

2. Estimateur du maximum de vraisemblance d'un convexe D sur base d'une réalisation d'un Processus de Poisson non homogène sur D

2.1. Vraisemblance du modèle

Considérons une réalisation $\overline{Y} = (Y_1, \ldots, Y_n)$ d'un tel processus. Celle-ci sera constituée de n points indépendants et identiquement distribués, de densité $f_D(x)$ égale à:

$$f_D(x) = \frac{q(x)\ 1_D(x)}{\int_D q(x)dx}$$

où q(x) est l'intensité inconnue du processus, supposée strictement positive.

Dès lors, la vraisemblance du vecteur $\overline{Y}$ sur D peut s'écrire:

$$f_D(\overline{Y}) = \prod_{i=1}^{n} f_D(Y_i)$$

$$= \prod_{i=1}^{n} \frac{q(Y_i)\ 1_D(Y_i)}{\int_D q(x)dx}$$

$$= \frac{1}{\left[\int_D q(x)dx\right]^n} \prod_{i=1}^{n} q(Y_i)\ 1_D(Y_i)$$

Puisque D est un domaine convexe; ceci implique que si l'on considère l'enveloppe convexe $H(\overline{Y})$ de notre réalisation, la vraisemblance se ré-écrit:

$$f_D(\overline{Y}) = \frac{1}{\left[\int_D q(x)dx\right]^n} \prod_{i=1}^{n} q(Y_i)\ 1_D(H(\overline{Y})) \quad (*)$$

...

2.2. Estimateur du convexe D

Comme pour le Processus de Poisson homogène, nous concluons de l'expression (*):

1) l'estimateur doit au moins contenir $H(\overline{Y})$ pour que la fonction indicatrice du domaine D soit non nulle;

2) l'estimateur ne peut pas être plus grand que $H(\overline{Y})$ car nous ajouterions une quantité positive au dénominateur et cela diminuerait la vraisemblance;

En conclusion, l'estimateur du maximum de vraisemblance du domaine convexe D sur base d'une réalisation d'un processus de Poisson non homogène sur D est l'enveloppe convexe $H(\overline{Y})$ de cette réalisation.

...

3. Analyse discriminante - PREMIER CRITERE: domaines disjoints

3.1. Fonction de densité en un point du domaine

Considérons $\bar{X} = (X_1, \ldots, X_n)$ la réalisation d'un Processus de Poisson non homogène d'intensité $q(x) > 0$ sur un domaine convexe D tel que:

$$D = \bigcup_{j=1}^{K} D_j$$

où les $(D_j)_{j \; j = 1 \ldots K}$ sont des domaines convexes et disjoints.

Tout individu de l'échantillon satisfait à une distribution dans D dont la densité $f_D(x)$ est égale à:

$$f_D(x) = \frac{q(x)\, 1_D(x)}{\int_D q(x)dx} \quad ;$$

Si $q_i(x)$ désigne la restriction de q à D_i c'est-à-dire $q_i(x) = q(x)\, 1_{D_i}(x)$, puisque D est l'union de K domaines disjoints D_j $(j = 1 \ldots K)$, nous pouvons écrire la densité f(x) sous la forme:

$$f(x) = \sum_{i=1}^{K} \frac{q_i(x)}{\int_D q(x)dx}$$

$$= \frac{\sum_{i=1}^{K} q_i(x)}{\int_D q(x)dx}$$

...

Posant: $\int_{D_i} q_i(x)dx$

$$p_i = \frac{\int_{D_i} q_i(x)dx}{\int_{D} q(x)dx} \quad \text{et}$$

$$f_i(x) = \frac{q_i(x)}{\int_{D_i} q(x)dx}$$

Nous obtenons une décomposition de la densité:

$$f(x) = \sum_{i=1}^{K} p_i f_i(x)$$

Remarques:

1) Comme, dans le cas du Processus de Poisson homogène, nous avons pu décomposer f(x) en une somme de $p_i f_i(x)$.
Il nous suffit, dès à présent, d'appliquer la règle d'affectation bayesienne:

"x sera affecté à la classe i" si $p_i f_i(x) \geqslant p_k f_k(x) \quad \forall k \neq i$

2) Les paramètres inconnus de la $i^{ème}$ population sont:
- le domaine convexe D_i lui-même;
- $q_i(x)$ i.e. la restriction de l'intensité inconnue $q(x)$ à D_i

Pour notre propos, rappelons que l'estimateur du maximum de vraisemblance du convexe D_i est l'enveloppe convexe des individus de ce groupe.

Quant au second paramètre inconnu, il faudra employer les estimations de densité qui sont le thème de la troisième partie de ce mémoire.

3) La règle d'affectation Bayesienne ne pourra s'appliquer que lorsque les enveloppes convexes $H(X_1), \ldots, H(X_K)$ sont disjointes, parce que nous avons supposé les domaines $D_1 \ldots D_K$ disjoints

...

3.2. Règle de décision

Soit $X_1 \ldots X_K$ les ensembles d'entrainement pour la population (1 ... K). La règle d'affectation bayesienne nous donne:

x est affecté à la classe i

$\Leftrightarrow$

$$p_i f_i(x) \geqslant p_k f_k(x) \qquad \forall k \neq i \qquad (1)$$

Développons cette relation:

$$\frac{q_i(x)}{\int_D q(x)dx} \geqslant \frac{q_k(x)}{\int_D q(x)dx} \qquad (2)$$

où $D = \bigcup_{i=1}^{K} D_i$

Comme les D_i sont disjoints, nous obtenons:

$$\sum_{i=1}^{K} \frac{q_i(x)}{\int_{D_i} q_i(x)dx} \geqslant \sum_{i=1}^{K} \frac{q_k(x)}{\int_{D_i} q_i(x)dx} \qquad (3)$$

...

3.2.1. PREMIER CAS: le point x_o, à classer, est à l'extérieur des enveloppes convexes $H(X_i)$ et $H(X_k)$

Si le point x_o est classé dans la $i^{ème}$ catégorie, nous estimons les domaines D_j de la manière suivante:

$$\hat{D}_j = H(X_i \bigcup (x_o)) \qquad j = i;$$

$$= H(X_j) \qquad j \neq i;$$

De même pour l'indice k, les domaines D_j seront estimés par

$$\bar{D}_j = H(X_k \bigcup (x_o)) \qquad j = k$$

$$= H(X_j) \qquad j \neq k$$

La règle bayesienne (2) devient:

$$\frac{q_i(x_o)}{\sum_{i=1}^{K} \int_{\hat{D}_i} q_i(x)dx} \geqslant \frac{q_k(x_o)}{\sum_{i=1}^{K} \int_{\bar{D}_i} q_i(x)dx} \qquad (4)$$

En particulier, si $q_i(x_o) = q_k(x_o)$, nous en déduisons

$$\sum_{i=1}^{K} \int_{\hat{D}_i} q_i(x)dx \leqslant \sum_{i=1}^{K} \int_{\bar{D}_i} q_i(x)dx$$

$$\sum_{j=1}^{K} \int_{H(X_j)} q(x)dx + \int_{H(X_i \bigcup (x_o))} q_i(x)dx - \int_{H(X_i)} q_i(x)dx \leqslant$$

$$\sum_{j=1}^{K} \int_{H(X_j)} q(x)dx + \int_{H(X_k \bigcup (x_o))} q_k(x)dx - \int_{H(X_k)} q_k(x)dx \qquad (5)$$

...

Notons par $S_i(x_o)$, la quantité suivante:

$$S_i(x_o) = \int_{H(X_i \bigcup (x_o))} q_i(x)dx - \int_{H(X_i)} q_i(x)dx$$

c'est-à-dire $S_i(x_o)$ représente l'intensité que x_o "ajoute" à l'enveloppe convexe de l'échantillon étiquetté du groupe i, lorsqu'il y est incorporé.

La nouvelle règle de classement devient:
"x_o est affecté à la classe i" si

$$S_i(x_o) \leqslant S_k(x_o) \quad \forall\, k \neq i$$

c'est-à-dire, le point x_o est affecté à la classe à laquelle l'intensité ajoutée par convexité est minimale.

...

3.2.2. SECOND CAS: le point x_o, à classer, appartient à une seule enveloppe convexe $H(X_i)$

D'après nos hypothèses, seul $q_i(x) = q(x)\ 1_{D_i}(x)$ sera strictement positif sur $\hat{D}_i = H(X_i)$.

Par conséquent, $\hat{p}_i f_i(x) > \hat{p}_k f_k(x)\ \forall\ k \neq i$ et donc le point x_o est affecté à la classe i.

Ainsi, nous rencontrons la première condition d'admissibilité de FISHER et VAN NESS:

"Si les enveloppes convexes $H(X_i)$ sont disjointes et x appartient à une de ces enveloppes convexes, alors x devra être affecté à la classe correspondante".

...

4. Analyse discriminante - SECOND CRITERE: domaines non disjoints

Nous nous intéressons au problème pour lequel les points, à classer, sont à l'intersection de plusieurs domaines convexes.

La règle d'affectation Bayesienne (1) ne change pas. Nous pouvons donc écrire l'inégalité (2):

$$\frac{q_i(x)}{\int_D q(x)dx} \geqslant \frac{q_k(x)}{\int_D q(x)dx} \qquad \forall\, k \neq i$$

Comme le point x appartient aux domaines convexes, il n'y a plus d'intensité ajoutée; ce qui implique que les dénominateurs restent les mêmes.

L'inégalité (2) devient:

$$q_i(x) \geqslant q_k(x) \qquad \forall\, k \neq i$$

C'est-à-dire, il faut affecter le point x à la classe pour laquelle l'intensité est maximale.

Remarque:

Notons la différence de méthode de classement proposée par SILVERMAN *.

Celui-ci attribue un poids de $\frac{1}{n}$ à chaque probabilité à priori p_i $(i = 1 \ldots K)$

Ceci implique que l'approche basée sur le maximum de vraisemblance affecterait le point x à la classe i si

$$p_i f_i(x) \geqslant p_k f_k(x) \qquad \forall\, k \neq i$$

$$\Leftrightarrow$$

$$\frac{1}{n} f_i(x) \geqslant \frac{1}{n} f_k(x) \qquad \forall\, k \neq i$$

$$f_i(x) \geqslant f_k(x) \qquad \forall\, k \neq i$$

Le critère de classement dépend, ici, des fonctions de densité de probabilité en un point relativement aux différents domaines.

(* se référer à SILVERMAN, Density Estimation, p° 121) ...

5. Conclusion

La généralisation de l'hypothèse du Processus de Poisson homogène en un Processus de Poisson non homogène nous permet, maintenant, de classer les points appartenant à l'intersection de plusieurs domaines convexes.

Cependant, il nous reste à estimer les intensités inconnues $q_i(x)$ des domaines $(D_i)_{i = 1 \ldots K}$

...

TROISIEME PARTIE: METHODES D'ESTIMATION DE DENSITE

1. Introduction

Dans le cadre du Processus de Poisson non homogène, les deux paramètres inconnus étaient:

a) les domaines convexes D_i qui ont été estimés par leur estimateur de maximum de vraisemblance, à savoir:

"l'enveloppe convexe de l'échantillon de la classe considérée."

b) les intensités inconnues $q_i(x)$ c'est-à-dire la restriction de l'intensité inconnue $q(x)$ au domaine convexe D_i.

L'objectif de cette partie est de montrer les différentes méthodes d'estimation de densité, car nous classerons tout individu selon la règle Bayesienne, en remplaçant les intensités inconnues par leur estimation suivant une des méthodes que nous allons décrire.

En pratique, dans le cadre de la nouvelle classification, les intensités seront estimées par le noyau de l'estimateur naïf de densité, appliqué au cas discret.

Nous appliquerons ce résulat à l'exemple de la télédétection qui sera l'objet de la quatrième partie.

...

2. Méthodes non paramétriques

Tout au long de ce chapitre, nous considérerons un échantillon $\underline{X}$ composé de n observations $X_1 \ldots X_n$.

Nous noterons par $\hat{f}(x)$ l'estimateur de densité proposé.

2.1. Les histogrammes

2.1.1. Définition

Nous définissons les fenêtres comme étant des intervalles $[x_o + mh, x_o + (m+1)h]$ pour des entiers m positifs et négatifs où

x_o : origine fixée

h: la largeur de la fenêtre

2.1.2. Estimateur

L'estimateur de densité déterminé par l'histogramme est défini par:

$$\hat{f}(x) = \frac{1}{nh}\left[\text{nombre d'observations } X_i \text{ appartenant à la même fenêtre que } x\right]$$

où h: largeur de la fenêtre définie à priori.

2.1.3. Remarque

Si nous supposons qu'il n'y a pas de subdivision, à priori, de la droite réelle en fenêtres c'est-à-dire que la subdivision dépendra des observations elles-mêmes; nous pouvons écrire l'estimateur comme:

$$f(x) = \frac{1}{n}\frac{\left[\text{nbre d'observations } x_i \text{ appartenant à la même fenêtre que } x\right]}{\left[\text{largeur de la fenêtre contenant } x\right]}$$

2.1.4. Utilité

Pour la présentation des données, les histogrammes représentent une classe extrêmement utile d'estimateurs, particulièrement dans le cas univarié. La présentation graphique des données bivariées et trivariées présentent plusieurs difficultés, il faudra envisager des estimateurs appropriés. ...

2.1.5. Exemple

Considérons la table comportant des données concernant les durées de traitement (en jours) de patients contrôlés dans une étude de suicide.

Table 2.1 *Lengths of treatment spells (in days) of control patients in suicide study.*

1	25	40	83	123	256
1	27	49	84	126	257
1	27	49	84	129	311
5	30	54	84	134	314
7	30	56	90	144	322
8	31	56	91	147	369
8	31	62	92	153	415
13	32	63	93	163	573
14	34	65	93	167	609
14	35	65	103	175	640
17	36	67	103	228	737
18	37	75	111	231	
21	38	76	112	235	
21	39	79	119	242	
22	39	82	122	256	

Nous pouvons considérer les histogrammes suivants, construits avec la même largeur de fenêtre, mais en différentes origines.

...

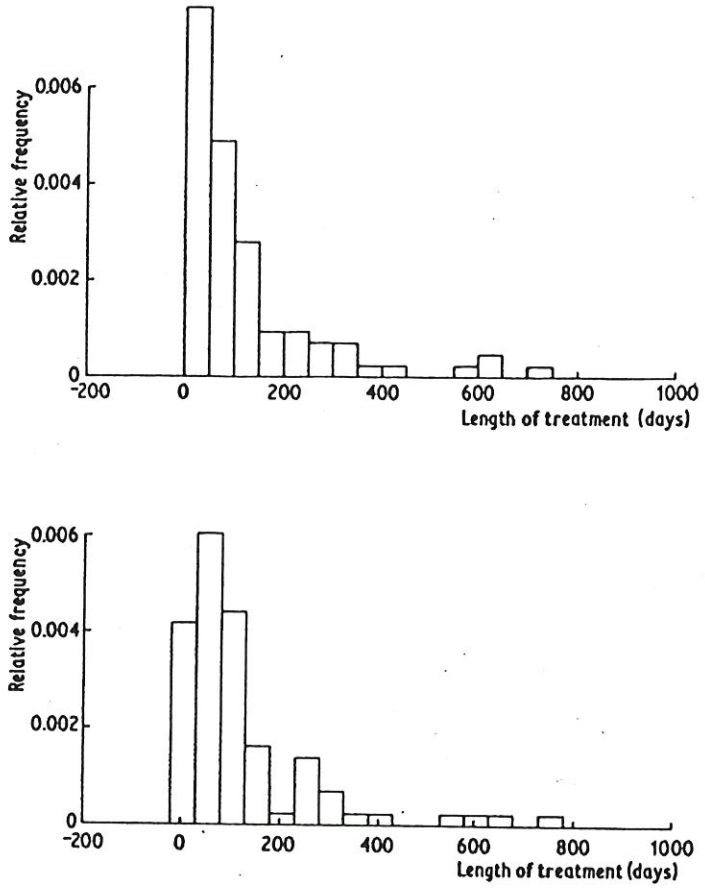


Fig. 2.1 *Histograms of lengths of treatment of control patients in suicide study.*

2.2. L'estimateur naïf

2.2.1. Construction

a) Si X est une variable aléatoire de densité de probabilité f, alors

$$f(x) = \lim_{h \to 0} \frac{1}{2h} P\left[x - h < X < x + h\right]$$

b) Pour n'importe quel h donné, nous pouvons estimer $P(x - h < X < x + h)$ par la proportion de l'échantillon qui tombe dans l'intervalle $(x - h, x + h)$.

Un estimateur $\hat{f}$ de densité peut être défini comme suit:

$$\hat{f}(x) = \frac{1}{2hn}\left[\text{nombre de } X_1, \ldots, X_n \text{ tombant dans l'intervalle } (x - h, x + h)\right]$$

Si nous définissons une fonction poids w spécifique, nous pouvons exprimer l'estimateur de la façon suivante.

2.2.2. Estimateur

L'estimateur naïf est donné par

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} w\left[\frac{x - X_i}{h}\right] \quad (*)$$

où la fonction poids w(x) est définie par

$$w(x) = \frac{1}{2} \quad \text{si } |x| < 1$$

$$= 0 \quad \text{sinon}$$

2.2.3. Interprétation

De la relation (*), nous constatons que l'estimateur naïf place une "boîte" de largeur 2 h et de poids (1 / 2 n h) sur chaque observation.

L'estimateur naïf, permet de construire un histogramme où chaque point est le centre d'un intervalle dont la largeur des fenêtres est de 2 h.

...

2.2.4. Exemple

Considérons la table comportant des données concernant les durées d'éruption (en minutes) de 107 éruptions du volcan "Old faithfull".

Table 2.2 *Eruption lengths (in minutes) of 107 eruptions of Old Faithful geyser.*

4.37	3.87	4.00	4.03	3.50	4.08	2.25
4.70	1.73	4.93	1.73	4.62	3.43	4.25
1.68	3.92	3.68	3.10	4.03	1.77	4.08
1.75	3.20	1.85	4.62	1.97	4.50	3.92
4.35	2.33	3.83	1.88	4.60	1.80	4.73
1.77	4.57	1.85	3.52	4.00	3.70	3.72
4.25	3.58	3.80	3.77	3.75	2.50	4.50
4.10	3.70	3.80	3.43	4.00	2.27	4.40
4.05	4.25	3.33	2.00	4.33	2.93	4.58
1.90	3.58	3.73	3.73	1.82	4.63	3.50
4.00	3.67	1.67	4.60	1.67	4.00	1.80
4.42	1.90	4.63	2.93	3.50	1.97	4.28
1.83	4.13	1.83	4.65	4.20	3.93	4.33
1.83	4.53	2.03	4.18	4.43	4.07	4.13
3.95	4.10	2.72	4.58	1.90	4.50	1.95
4.83	4.12					

...

L'estimateur naïf construit sur base de ces données, h = 0.25 est le suivant:

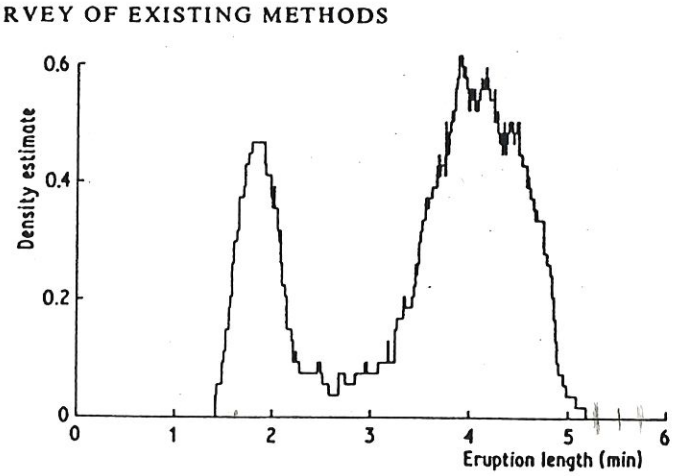


Fig. 2.2. *Naive estimate constructed from Old Faithful geyser data, h = 0.25.*

1.67 4.93
- 0.25 0.25
1.42 5.18

...

2.3. L'estimateur des noyaux

2.3.1. Estimateur

Par analogie avec l'estimateur naïf,

$$\hat{f}(x) = \frac{1}{n\,h} \sum_{i=1}^{n} K\left[\frac{x - X_i}{h}\right]$$

où . le noyau K est une fonction de densité de probabilité symétrique telle que:

a) K est partout non négatif

b) $\int_{-\infty}^{+\infty} K(x)dx = 1$

. h est la largeur de fenêtre, appelée "smoothing parameter".

2.3.2. Interprétation

L'estimateur des noyaux est une somme de "bosses" placées sur chaque observation.

La fonction noyau K détermine la forme des "bosses" tandis que h détermine leur largeur.

Les figures suivantes illustrent cet estimateur pour différentes valeurs de h.

...

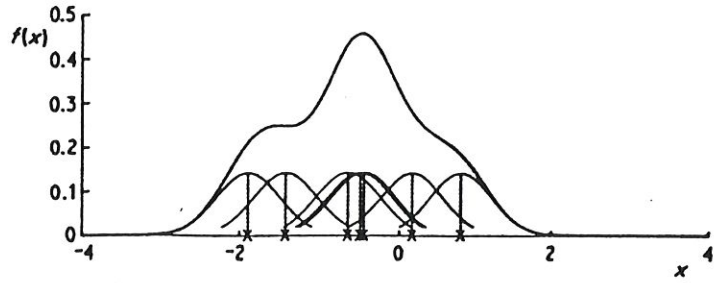


Fig. 23.1 *Kernel estimate showing individual kernels. Window width 0.4.*

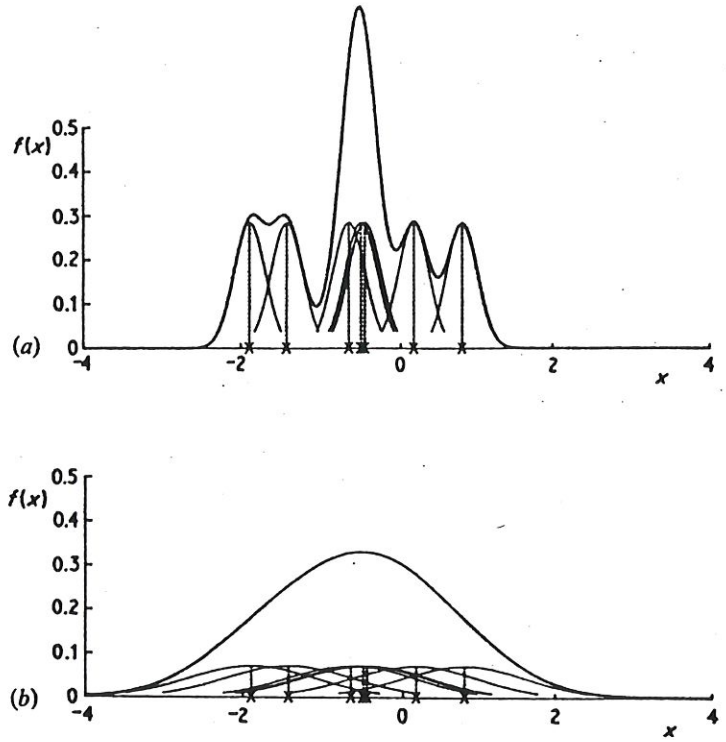


Fig. 23.2 *Kernel estimates showing individual kernels. Window widths: (a) 0.2; (b) 0.8.*

Notons que lorsque h augmente, la fonction devient plus "douce".

...

2.3.3. Exemple

La figure suivante montre l'estimateur des noyaux construits sur les données concernant les temps d'éruptions des 107 éruptions du volcan "Old faithfull".

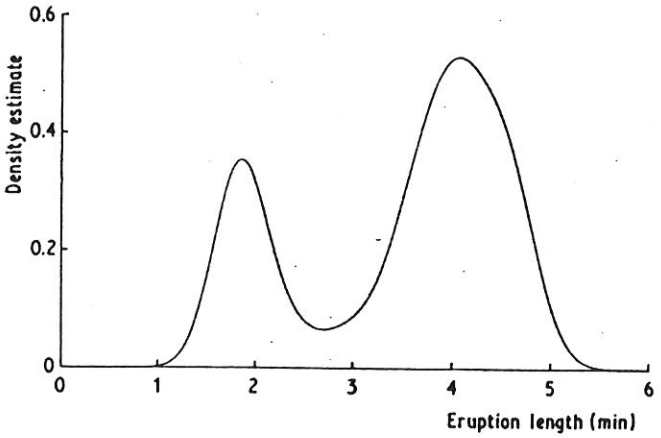


Fig. 2.33 *Kernel estimate for Old Faithful geyser data, window width 0.25.*

...

2.4. L'estimateur de densité du $k^{ième}$ plus proche voisin:

2.4.1. Estimateur

Considérons la distance euclidienne entre deux points:

$$d(x,y) = |x - y|$$

Faisons l'hypothèse suivante:

$$\forall\, t \quad d_1(t) \leqslant d_2(t) \leqslant \ldots \leqslant d_n(t)$$

où les $d_i(t)$ sont les distances ordonnées, du point t aux points de l'échantillon.

L'estimateur de densité du $k^{ième}$ plus proche voisin est défini par

$$\hat{f}(t) = \frac{k - 1}{2\, n\, d_k(t)}$$

2.4.2. Interprétation

Par analogie avec la construction de l'estimateur naïf, si la densité au point t est f(t), on s'attend à ce que $2\, r\, n\, \hat{f}(t)$ observations tombent dans l'intervalle $(t - r, t + r)\ \forall\ r>o$

En effet,

$$\hat{f}(t) = \frac{1}{2rn} \left[\text{nombre d'observations tombant dans l'intervalle } (t - r, t + r) \right]$$

Par conséquent, on s'attend à ce que $(k - 1)$ observations tombent dans l'intervalle $(t - d_k(t),\ t + d_k(t))$, pour le cas de l'estimateur du $k^{ième}$ plus proche voisin.

...

2.4.3. Exemple

La figure suivante représente l'estimateur du $k^{ième}$---- plus proche voisin pour les données concernant les éruptions ($k = 20$).

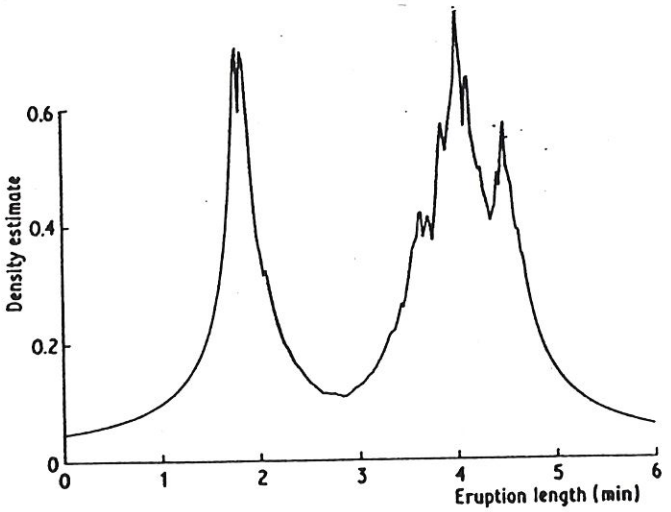


Fig. 2.4 *Nearest neighbour estimate for Old Faithful geyser data, $k = 20$.*

2.4.4. Généralisation

L'estimateur généralisé du $k^{ième}$---- plus proche voisin est défini par:

$$\hat{f}(t) = \frac{1}{n\, d_k(t)} \sum_{i=1}^{n} K\left[\frac{t - X_i}{d_k(t)}\right]$$

$\hat{f}(t)$ est l'estimateur du noyau évalué en t avec une largeur de fenêtre correspondant à $d_k(t)$.

Ici, le degré de "régularité" est contrôlé par le choix de l'entier k, supposé considérablement plus petit que la taille n de l'échantillon.

...

2.5. Méthode du noyau variable

2.5.1. Estimateur

Considérons K une fonction de noyau et k un entier positif.

L'estimateur du noyau variable possédant h comme largeur de bande est défini par

$$\hat{f}(t) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h\, d_{i,k}} K\left[\frac{t - X_i}{h\, d_{i,k}}\right]$$

où $d_{i,k}$: distance de X_i au $k^{\text{ième}}$ point le plus proche dans l'ensemble comprenant les autres (n - 1) observations.

2.5.2. Interprétation

L'estimateur est construit de la même façon que l'estimateur du noyau, mais le paramètre d'échelle des "bosses" placées aux données peut varier d'un point à l'autre; la largeur de la fenêtre placée au point X_i sera proportionnelle à $d_{i,k}$.

...

2.5.3. Exemple

Le schéma suivant illustre l'estimateur du noyau variable pour les données concernant l'étude de suicide.

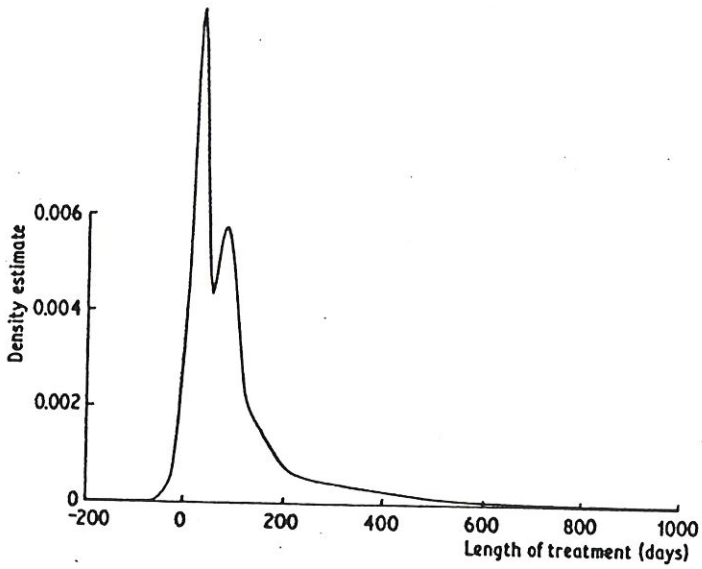


Fig. 2.5 *Variable kernel estimate for suicide study data, $k=8$, $h=5$.*

...

2.6. Méthode du noyau adaptatif

2.6.1. Estimateur

PAS 1: Trouver un estimateur pilote $\tilde{f}(t)$ qui satisfait à la condition:

$$\tilde{f}(X_i) > 0 \quad \forall i$$

PAS 2: Définir des facteurs de largeur de bandes locales λ_i tel que

$$\lambda_i = (\tilde{f}(X_i) / g)^{-\alpha}$$

où . g est la moyenne géométrique de $\tilde{f}(X_i)$ c'est-à-dire

$$\log g = \frac{1}{n} \sum_{i=1}^{n} \log \tilde{f}(X_i)$$

. α est le paramètre de sensibilité tel que $0 \ll \alpha \ll 1$

PAS 3: Définir l'estimateur de noyau adaptatif $\hat{f}(t)$ par

$$\hat{f}(t) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{(h \lambda_i)^d} K\left[\frac{t - X_i}{\lambda_i h}\right]$$

où . K est la fonction noyau et

. h est la largeur de bande

2.6.2. Interprétation

Cet estimateur combine à la fois l'approche du plus proche voisin et celle du noyau. On construit un estimateur de noyau composé de "bosses" ou de noyaux placés aux observations, mais la largeur de fenêtre des noyaux peut varier d'un point à l'autre.

...

2.6.3. Exemple

La figure suivante représente l'estimateur de noyau adaptatif pour les données concernant l'étude de suicide.

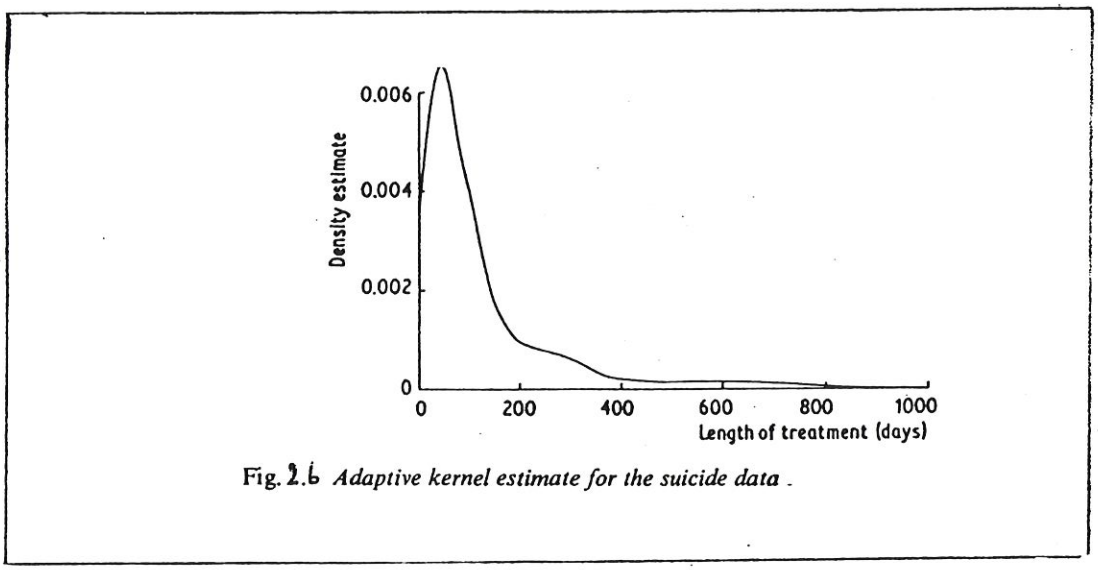


Fig. 2.6 *Adaptive kernel estimate for the suicide data .*

...

2.7. Estimateurs de fonction de poids général

2.7.1. Estimateur

L'estimateur est défini par

$$\hat{f}(t) = \frac{1}{n} \sum_{i=1}^{n} w(X_i, t) \qquad (*)$$

où $w(x,y)$ est une fonction à deux arguments telle que:

- $\int_{-\infty}^{+\infty} w(x,y)\, dy = 1$
- $w(x,y) \geq 0 \quad \forall x \quad \forall y$

2.7.2. Utilité

1) C'est un concept qui rassemble des estimateurs déjà cités.

   En posant,

   $w(x,y) = \frac{1}{h(x)}$ si x et y appartiennent à la même fenêtre

   $= 0$ sinon

   Nous obtenons le cas de l'histogramme.

2) Cet estimateur permet de définir d'autres estimateurs que nous n'avons pas énoncés et qui sont de la forme (*).

...

3. Fonction noyau K(x)

o Nous allons donner une liste des noyaux K(x) utilisés dans l'estimation de densités. En pratique, nous estimerons les intensités inconnues, dans le cadre du Processus de Poisson non homogène, par le noyau de l'estimateur naïf.

Liste des noyaux
****

NOYAU	K(x)
- Rectangulaire (ou de l'estimateur naïf)	$\frac{1}{2}$ si $\|x\| < 1$; o sinon
- Triangulaire	$1 - \|x\|$ si $\|x\| < 1$; o sinon
- A double poids	$\frac{15}{16}(1 - x^2)^2$ si $\|x\| < 1$; o sinon
- Epanechnikov	$\frac{3}{4}(1 - \frac{1}{5}k^2)/\sqrt{5}$ si $\|x\| < \sqrt{5}$; o sinon

Note:

Ces noyaux sont définis dans le cas continu sur la droite réelle.

...

4. Le noyau de l'estimateur naïf

4.1. Formalisation du problème

. Nous voulons estimer les intensités inconnues par le noyau de l'estimateur naïf appliqué au cas discret.

L'estimateur naïf de densité se définit alors par:

$$\hat{f}(x) = \frac{1}{n} \sum_{j=1}^{K} K\left[x - X_j\right]$$

où . $K(x) = \dfrac{1}{(2h+1)^d}$ si $|x| \leqslant h$

$= 0$ sinon

. h est la largeur de bande

. d est la dimension de l'espace dans lequel on travaille.

. Remarquons que K(x) définit une fonction de densité de probabilité symétrique telle que

1°) $\sum_{x=-h}^{h} K(x) = 1$

2°) K(x) est partout non négatif

. Notre problème réside dans le choix du paramètre h qui fixe la largeur de fenêtre qui sera placée en chaque observation X_i .

Nous allons aborder la recherche du h optimum par l'approche de SILVERMAN * appliquée au cas discret.

* se référer à B.W. SILVERMAN, Density Estimation, p° 52

...

4.2. Approche de B.W. SILVERMAN

4.2.1. Principe de départ

* SILVERMAN utilise une approche originale de la vraisemblance. Il part du principe suivant: Supposons qu'à un ensemble de n données, nous disposons en plus d'une nouvelle observation indépendante y.

Alors le log-vraisemblance de f, où f est la densité sous-jacente à l'observation Y, serait log f(Y).

Soit $\hat{f}$ l'estimateur de la densité, dépendant d'une largeur de fenêtre h où les observations $X_1, \ldots, X_n$ sont fixées.

Le log-vraisemblance sera estimé par $\log \hat{f}(Y)$ où cette fonction dépendra de h.

Si maintenant, nous supposons que l'on n'a pas l'observation indépendante Y, nous pourrions retirer une observation X_i de l'ensemble initial des données; où X_i est n'importe quelle observation de l'échantillon.

X_i jouerait le rôle de l'observation indépendante Y, et nous obtiendrons une fonction de log-vraisemblance

$$CV(h) = \frac{1}{n} \sum_{i=1}^{n} \log \hat{f}_{-i}(X_i)$$

où $\hat{f}_{-i}(X_i)$ est l'estimateur de densité construit à partir des données $X_1 \ldots X_n$ excepté X_i, c'est-à-dire:

$$\hat{f}^{i}_{-i}(X_i) = \frac{1}{n-1} \sum_{j \neq i} K(X_i - X_j) \qquad (*)$$

...

4.2.2. Choix du paramètre h

Le choix du paramètre h se réalisera comme suit:

le h optimum est celui qui maximisera le fonction de log-vraisemblance CV(h)

4.2.3. Bornes du paramètre h

4.2.3.1. Condition de recouvrement d'un point

Selon les hypothèses de SILVERMAN, si nous voulons éviter le cas où la fonction CV(h) serait égale à $(-\infty)$, il faudrait que:

$$\forall i = 1 \ldots n \quad \hat{f}_{-i}(X_i) \neq 0;$$

Développons cette condition:

$$\forall i = 1 \ldots n \ \exists j \ (1 \ldots n) \quad i \neq j \quad t\,q \quad K\left[X_i - X_j\right] \neq 0;$$

Pour ce faire, il faut que:

$$\forall i = 1 \ldots n \ \exists j \ (1 \ldots n) \quad i \neq j \quad t\,q \quad |X_i - X_j| \leqslant h;$$

où $|X_i - X_j|$ est la maxi-distance entre les points X_i et X_j

c'est-à-dire,

$$|X_i - X_j| = \max_{k \in (1 \ldots d)} (|X_{ii} - X_{j1}|, \ldots, |X_{id} - X_{jd}|)$$

où d est la dimension de l'espace de travail.

<u>Définition</u>

Nous dirons que le point X_i recouvre le point X_j si la distance considérée $d(X_i, X_j) \leqslant h$

Dans notre cas, un point X_i, sera recouvert par le point X_j si $|X_i - X_j| \leqslant h$

Ou encore

$$\max_{k \in (1 \ldots d)} (|X_{ii} - X_{j1}|, \ldots, |X_{id} - X_{jd}|) \leqslant h$$

...

4.2.3.2. Borne inférieure pour h: h_{min}

Nous disposons donc de n observations $X_1, \ldots, X_n$

. Pour qu'un point X_i soit recouvert par au moins une observation X_j ($j \in (1, \ldots, n), j \neq i$), il faut que:

$$\min_{\substack{j \in (1 \ldots n) \\ j \neq i}} |X_i - X_j| \leqslant h \quad (1)$$

Cette condition assure que le point X_i est recouvert par au moins un point qui est son plus proche voisin au sens de la distance considérée.

. N'oublions pas qu'il faut la condition (1) pour tous les points X_i ($i = 1 \ldots n$) ceci implique que la condition (1) devient

$$\max_{i \in (1 \ldots n)} \min_{\substack{j \in (1 \ldots n) \\ j \neq i}} |X_i - X_j| \leqslant h \quad (2)$$

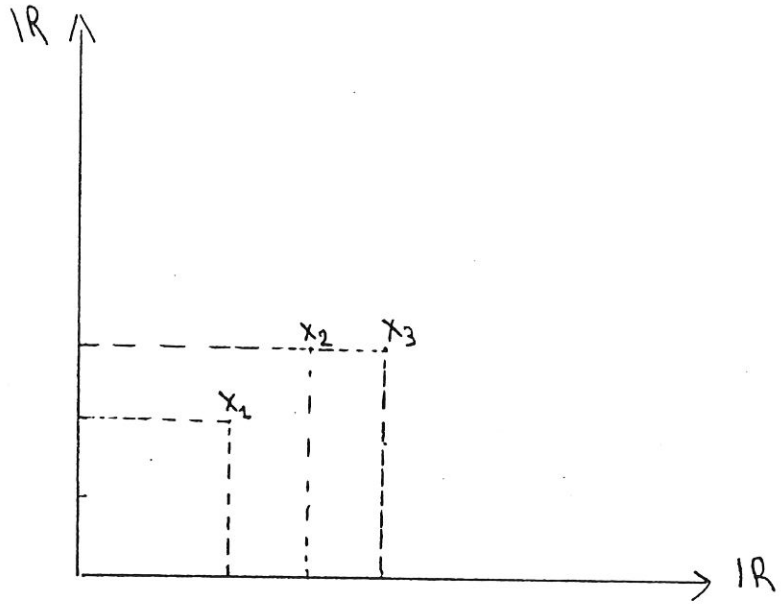
En conclusion, pour éviter le cas où un point X_i, ne serait pas recouvert, il faut que h soit plus grand ou égal à la plus grande des plus petites maxi distances, ou encore

$$h \geqslant \max_{i \in (1 \ldots n)} \min_{\substack{j \in (1 \ldots n) \\ j \neq i}} \max_{k \in (1 \ldots d)} (|X_{i1} - X_{j1}|, \ldots, |X_{id} - X_{jd}|) \overset{not}{=} h_{min}$$

...

Illustration

Nous disposons de trois observations dans le plan: n = 3,d = 2



1. Calcul des maxi-distances pour chaque point

pour le point x_1 : $|x_1 - x_2| = 1$; $|x_1 - x_3| = 2$

pour le point x_2 : $|x_2 - x_1| = 1$; $|x_2 - x_3| = 1$

pour le point x_3 : $|x_3 - x_1| = 2$; $|x_3 - x_2| = 1$

2. Prendre le min des maxi-distances en chaque point

pour le point x_1 : $h \geqslant 1$

pour le point x_2 : $h \geqslant 1$

pour le point x_3 : $h \geqslant 1$

3. Tous les points doivent être recouverts par au moins un point

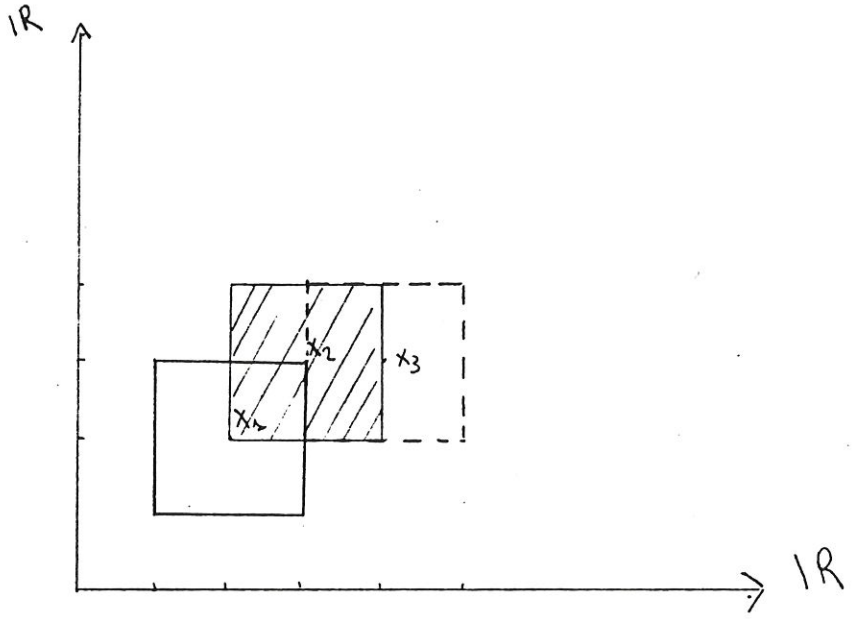
$h \geqslant$ Max (1, 1, 1) , c'est-à-dire

$h \geqslant 1 = h_{min}$

...

4. Vérification: $h_{min} = 1$

Nous voyons sur la figure suivante, que tous les points sont recouverts par au moins un point qui est leur plus proche voisin



le point x_1 est recouvert par x_2

le point x_2 est recouvert par x_1 et x_3

le point x_3 est recouvert par x_2

...

4.2.3.3. Borne supérieure pour h: h_{max}

Nous disposons de n observations
$X_1 \ldots X_n$

. Un point X_i est recouvert par un point X_j ($j \in (1...n)$, $j \neq i$), si par définition $|X_i - X_j| \leqslant h$

Si nous désirons qu'un point X_i soit recouvert par tous les X_j ($j = 1...n$ $j \neq i$), c'est-à-dire que le nombre maximum d'observations recouvrant X_i soit n - 1, il faut

que $h \geqslant \underset{\substack{j \in (1...n) \\ j \neq i}}{\text{Max}} |X_i - X_j|$

Notons que si $h \geqslant \underset{\substack{j \in (1...n) \\ j \neq i}}{\text{Max}} |X_i - X_j|$,

ceci implique que le numérateur de la relation (*) est toujours égal à (n - 1), mais le h qui apparaît au dénominateur augmente, dès lors $\hat{f}_{-i}(X_i)$ diminue et par conséquent la fonction log-vraisemblance CV(h) à maximiser décroit.

. Pour chaque observation X_i, le h peut donc varier jusqu'au Maximum des maxi-distances.

Comme il nous faut cette condition pour tous les X_i ($i = 1...n$), celle-ci sera vérifiée lorsque:

$$h \leqslant \underset{i \in (1...n)}{\text{Max}} \; \underset{\substack{j \in (1...n) \\ j \neq i}}{\text{Max}} |X_i - X_j| = h_{max} \qquad (3)$$

...

En conclusion,

D'après les relations (2) et (3), le h optimum sera tel que

$$h \in \left[\max_{i \in (1\ldots n)} \min_{\substack{j \in (1\ldots n) \\ j \neq i}} |X_i - X_j| ; \max_{i \in (1\ldots n)} \max_{\substack{j \in (1\ldots n) \\ j \neq i}} |X_i - X_j| \right]$$

où

$$|X_i - X_j| = \max_{k \in (1\ldots d)} (|X_{i1} - X_{j1}|, \ldots, |X_{id} - X_{jd}|)$$

avec d: la dimension de l'espace de travail

C'est-à-dire, h doit être compris entre la plus grande des plus petites maxi-distances et la plus grande des plus grandes maxi-distances.

Illustration

Reprenons l'exemple de la page 57

1. Calcul du Max des maxi-distances pour chaque point

Pour le point x_1 : $h \leqslant 2$

Pour le point x_2 : $h \leqslant 1$

Pour le point x_3 : $h \leqslant 2$

2. Calcul de la borne supérieure

$h \leqslant$ Max (2, 1, 2), c'est-à-dire

$h \leqslant 2 = h_{max}$

ceci implique que le h optimum est tel que: $h \in [1,2]$

Recherche du h optimum

Nous noterons par a_i le nombre de points qui recouvrent le point X_i ;

Nous calculerons, pour chaque h, la vraisemblance, où nous omettrons n: la taille de l'échantillon qui est constante à chacun de nos calculs.

...

h = 1

. $a_1 = 1 \quad a_2 = 2 \quad a_3 = 1$

La vraisemblance s'écrit:

$$\prod_{i=1}^{3} \hat{f}_{-i}(X_i) = \frac{1}{(2h+1)^2} \cdot \frac{2}{(2h+1)^2} \cdot \frac{1}{(2h+1)^2}$$

$$= \frac{2}{729}$$

$$= 0.0027$$

h = 2

. $a_1 = 2 \quad a_2 = 2 \quad a_3 = 2$

La vraisemblance s'écrit:

$$\prod_{i=1}^{3} \hat{f}_{-i}(X_i) = \frac{2}{(2h+1)^2} \cdot \frac{2}{(2h+1)^2} \cdot \frac{2}{(2h+1)^2}$$

$$= \frac{8}{15625}$$

$$= 0.0005$$

Le h optimum est atteint en $h = h_{min} = 1$, c'est-à-dire

le h optimum est atteint quand h est le Max Min des maxi-distances.

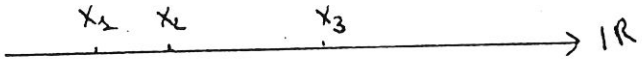
...

4.2.3.4. Remarques

1. Dans la plupart des exemples pratiques; le h optimum correspond souvent au h_{min}, c'est-à-dire au Max Min des maxi-distances, mais ce n'est pas toujours le cas:

   Voici un contre-exemple:

   Nous disposons de trois observations sur la droite réelle: n = 3, r = 1



1) Calcul des maxi-distances pour chaque point

pour le point x_1 : $|x_1 - x_2| = 1$, $|x_1 - x_3| = 3$

pour le point x_2 : $|x_2 - x_1| = 1$, $|x_2 - x_3| = 2$

pour le point x_3 : $|x_3 - x_1| = 3$, $|x_3 - x_2| = 2$

2) Recherche du h_{min}

a) Calcul de la borne inférieure pour chaque point

pour le point x_1 ; $h_{min}^{(1)} = 1$

pour le point x_2 ; $h_{min}^{(2)} = 1$

pour le point x_3 ; $h_{min}^{(3)} = 2$

b) Calcul du h_{min} pour tous les points

$$h_{min} = \underset{i \in (1, \ldots, 3)}{\text{Max}} h_{min}^{(i)} = 2$$

...

3) Recherche du h_{max}

a) Calcul de la borne supérieure pour chaque point

pour le point x_1 ; $h_{max}^{(1)} = 3$

pour le point x_2 ; $h_{max}^{(2)} = 2$

pour le point x_3 ; $h_{max}^{(3)} = 3$

b) Calcul du h_{max} pour tous les points

$$h_{max} = \underset{i \in (1, \ldots, 3)}{\text{Max}} \; h_{max}^{(i)} = 3$$

4) Recherche du h optimum

h optimum $\in [2,3]$

h = 2

$a_1 = 1$; $a_2 = 2$; $a_3 = 1$

La vraisemblance $\prod_{i=1}^{3} \hat{f}_{-i}(x_i) = \frac{1}{2h+1} \frac{2}{2h+1} \frac{1}{2h+1}$

$$= \frac{2}{125}$$

$$= 0.016$$

...

h = 3

$a_1 = 2 \quad a_2 = 2 \quad a_3 = 2$

La vraisemblance $\prod_{i=1}^{3} \hat{f}_{-i}(X_i) = \frac{2}{2h+1} \frac{2}{2h+1} \frac{2}{2h+1}$

$$= \frac{8}{343}$$

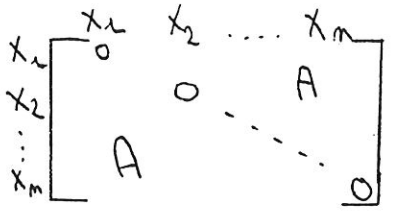
$$= 0.0204$$

Le h optimal est atteint pour $h = h_{max} = 3$

Note: Pour faciliter les calculs, nous avons omis la constante n

...

2. D'un point de vue pratique, si nous considérons X, la matrice des maxi-distances:

$X =$ 

Nous constatons que X est une matrice symétrique, diagonale.

. h doit varier entre d'une part la plus grande des plus petites maxi-distances sur les lignes de X et d'autre part la plus grande des plus grandes maxi-distances sur les lignes de X.
Nous pouvons déjà avoir h_{min} et h_{max}.

. Nous savons que le point x_i recouvre le point x_j ssi $d(x_i, x_j) \leqslant h$; ceci implique que le nombre de points qui recouvre x_i est le nombre de distances inférieures ou égales à h sur la ligne i de X

$a_i = \#$ de distances $\leqslant h$ sur la ligne i de X

La connaissance de la matrice X, nous permet de retrouver le h optimal qui maximise l'expression suivante:

$$\log \prod_{i=1}^{n} \hat{f}_{-i}(X_i) = \log \prod_{i=1}^{n} \frac{a_i}{(2h+1)^d}$$

$$= -\, n\, d\, \ln (2h+1) + \sum_{i=1}^{n} \ln a_i$$

où n est la taille de l'échantillon

d est la dimension de l'espace de travail

...

Principe général:

1. Calcul de la matrice X pour chaque classe

2. Recherche de:

   h_{min} : plus grande des plus petites maxi-distances sur les lignes de X

   h_{max} : plus grande des plus grandes maxi-distances sur les lignes de X

3. Pas itératif

   1er PAS: $h^{(o)} = h_{(min)}$

   . calcul de a_i = #de maxi-distances $\leqslant h^{(o)}$ sur la ligne i de X

   . calcul du log-vraisemblance pour $h^{(o)}$

   $$CV(h^{(o)}) = \left[- n \ln (2h^{(o)} + 1)^d + \sum_{i=1}^{n} \ln a_i \right]$$

   2e PAS: $h^{(o)} = h_{(min)} + 1$

   . calcul du a_i correspondant

   . calcul du $CV(h^{(1)})$ correspondant

   .
   .
   .

   k---- pas: $h^{(k)} = h_{max}$

   . calcul des a_i = #de maxi-distances $\leqslant h^{(k)}$

   . $$CV(h^{(k)}) = \left[- n \ln (2h^{(i)} + 1)^d + \sum_{i=1}^{n} \ln a_i \right]$$

4. Choix du h optimum

   $h_{opt.}$ est le $h^{(i)}$ qui maximise la fonction $CV(h^{(i)})$

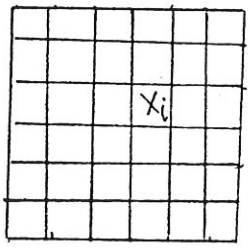
...

4.3. Problème important

L'approche proposée par SILVERMAN suppose qu'un point ne peut pas être recouvert par lui-même.

Considérons le cas discret suivant:

Soit la grille représentée ci-dessous:



Celle-ci est constituée de petits carrés pouvant contenir 0 ou 1 observation.

Supposons que chaque point peut être recouvert par lui-même.

Nous voulons retrouver le h qui maximise la fonction log-vraisemblance suivante:

$$\text{log-vraisemblance} = -n \ln (2h+1)^2 + \sum_{i=1}^{n} \ln a_i$$

où a_i est le nombre de points recouvrant X_i

Analysons le log-vraisemblance pour les valeurs suivantes de h:

...

$$\text{Pour } h^{(1)} = o$$

Le point X_i est recouvert par lui-même; ceci implique que:

$a_i = 1 \quad \forall\, i = 1 \ldots n$ et

le log-vraisemblance $(h^{(1)}) = - n \ln 1 + n \ln 1 = o$

$$\text{Pour } h^{(2)} \neq o$$

Puisque chaque case peut comporter zéro ou une observation, le nombre de points qui recouvrent X_i est plus petit ou égal au nombre de cases appartenant à la surface du carré de côté $2h + 1$

$$a_i \leqslant (2h + 1)^2 \qquad \forall i = 1 \ldots n$$

Dès lors, la fonction log-vraisemblance CV(h) est telle que le

$$\text{log-vraisemblance } (h^{(2)}) \leqslant - n \ln (2h + 1)^2 + n \ln a_i$$

$$\leqslant - n \ln (2h + 1)^2 + n \ln (2h + 1)^2$$

$$\leqslant o$$

Par conséquent, le h optimum est atteint pour $h^{(1)} = o$ c'est-à-dire dans la situation où chaque point n'est recouvert que par lui-même.

...

* Nous pouvons généraliser le cas de la grille dans le plan à un hypervolume dans un espace à d dimensions. L'hypervolume sera constitué de petits hypercubes pouvant contenir 0 ou 1 observation.
Nous voulons retrouver le h qui maximise la fonction log-vraisemblance.

$$\text{log-vaisemblance}\ (h) = -\ n\ \ln\ (2h+1)^{d} + \sum_{i=1}^{n} \ln\ a_i$$

Pour $h^{(1)} = o$

Le point X_i est recouvert par lui-même, ceci implique que:

$a_i = 1 \quad \forall\ i = 1 \ldots n$ et

$$\text{log-vaisemblance}\ (h^{(1)}) = -\ n\ \ln\ 1 + n\ \ln\ 1 = o$$

Pour $h^{(2)} \neq o$

Puisque chaque hypercube peut comporter zéro ou une observation, le nombre maximum de points recouvrant X_i est le nombre d'hypercubes appartenant à l'hypervolume de côté 2h + 1

$$a_i \leqslant (2h+1)^{d} \quad \forall\ i = 1 \ldots n$$

Dès lors, la fonction log-vraisemblance CV(h) est telle que:

$$\text{log-vraisemblance}\ (h^{(2)}) \leqslant -\ n\ \ln\ (2h+1)^{d} + n\ \ln\ a$$

$$\leqslant -\ n\ \ln\ (2h+1)^{d} + n\ \ln\ (2h+1)^{i}$$

$$\leqslant o$$

Par conséquent, le h optimum est atteint pour $h^{(1)} = o$.

...

En conclusion:

Si nous n'utilisons pas l'hypothèse de SILVERMAN qui suppose que chaque point ne peut pas être recouvert par lui-même, nous obtenons que le h optimum qui maximise la fonction de log-vraisemblance est toujours égal à zéro, dans n'importe quelle situation.
La vraisemblance serait maximale lorsque chaque point n'est recouvert que par lui-même.

...

QUATRIEME PARTIE: EVALUATION DE LA NOUVELLE METHODE DE CLASSIFICATION POUR LE TRAITEMENT D'IMAGES EN TELEDETECTION

1. Introduction

Dans le contexte des activités liées à la gestion du territoire, la télédétection constitue une source d'informations multiples.

Citons, par exemple, l'humidité des sols, leur occupation, la biomasse présente, etc.

Il y a encore peu, notamment au début du siècle, pour réaliser ce type de projet, l'être humain était obligé de mettre en oeuvre tout un capital humain pour arriver à ses fins.

En effet, la seule solution était le recensement systématique sur le terrain. La télédétection ne remplace pour autant pas les méthodes traditionnelles d'acquisition de ces données, mais en relation avec ces dernières, elle permet de compléter les résultats déjà existants.

Ces informations sont disponibles grâce à un satellite qui enregistre dans différents canaux spectraux les intensités lumineuses réfléchies par le sol, et crée ainsi une image numérique d'une région. Ces images numériques sont susceptibles d'être exploitées par des moyens informatiques.

Le traitement de l'image qui nous intéresse plus particulièrement est la classification des pixels, afin de déterminer l'occupation du sol.

Nous proposons donc la méthode de classification basée sur le Processus de Poisson non homogène.

Le critère de ressemblance recourt à l'intensité du processus et à la notion d'enveloppe convexe.

...

2. Principe du traitement d'image par classification

De façon générale, une classification est la distribution systématique d'objets en diverses "catégories", d'après des "critères" précis.

2.1. Définition des classes

La définition des classes est envisageable avant ou après l'opération de classification, il s'agit alors d'une "classification supervisée" respectivement "non supervisée".

. Classification supervisée

On définit un certain nombre de classes thématiques, par exemple, la forêt, l'eau, ... Le traitement de l'image consiste à reconnaître les pixels qui présentent une signature spectrale correspondant à une des classes sélectionnées.

. Classification non supervisée

Cette fois, les pixels sont regroupés selon la ressemblance de leur signature spectrale sans référence à d'éventuelles classes thématiques. Lorsque l'opération de classification est finie, les différents groupes de pixels apparus sont associés à une classe, d'après la nature des signatures inclues dans ce groupe.

...

2.2. Critère

L'attribution d'un pixel à une classe va dépendre de la ressemblance entre sa signature et celle caractéristique de la classe candidate. Le critère de classification est donc un critère de ressemblance. Suivant la nature de ce critère, on distingue d'une part les méthodes déterministes, et d'autre part les méthodes statistiques.

. Méthode déterministe

Dans le cas le plus courant, le critère de ressemblance entre deux pixels est la distance euclidienne. Plus cette distance est faible, plus leur ressemblance est grande.

. Méthode statistique

On suppose que la signature spectrale de chaque pixel est un vecteur aléatoire dans IR^n, n est le nombre de bandes spectrales, avec une certaine densité de probabilité. Le critère de ressemblance d'un pixel à une classe est mesuré par le risque de mauvais classement lorsqu'on affecte ce pixel à cette classe *. Le calcul de ce risque passe par la connaissance de la fonction de densité de probabilité.

Une méthode est dite paramétrique, si on impose que cette fonction dépende d'un paramètre, par exemple si on veut que la fonction de probabilité soit gaussienne.

Dans le cas contraire, il s'agit d'une méthode non-paramétrique.

* se référer au point 2.4.2. Risque bayesien.

...

2.3. Formalisation du problème

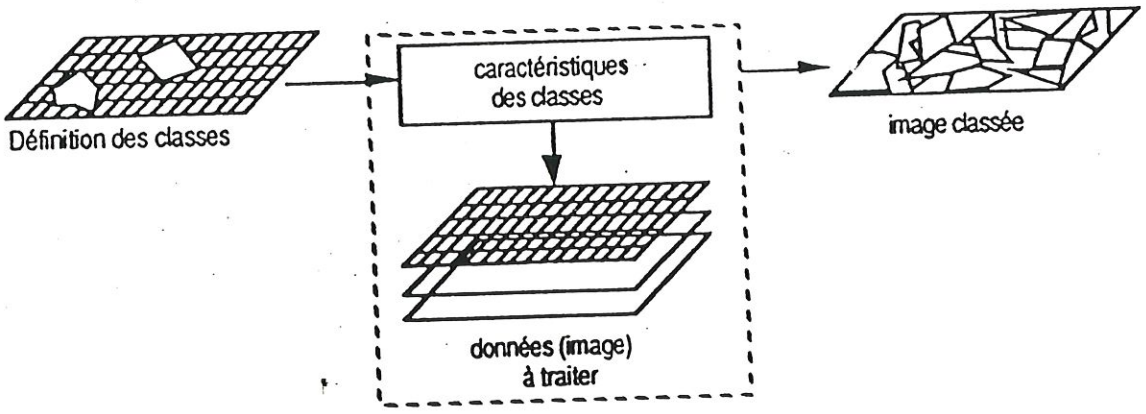
Notre propos est d'étudier une méthode statistique non-paramétrique de la classification supervisée.

Pour ce faire, nous disposons de classes de références.

Ensuite, tous les points devant former la carte, sont classés par rapport à ces références. Notre but, obtenir un produit fini, la carte.

La figure suivante illustre le principe général. La classification se déroule en trois étapes.

D'abord, les classes à identifier sont définies en choisissant pour chacune des zones de calage. On extrait de ces dernières les caractéristiques spectrales des classes, puis on procède à la classification d'image.



Principe général du traitement d'image par classification supervisée (tiré de [Abednego, 1989] p.90)

...

3. Problématique du traitement d'image par classification

Nous pouvons rencontrer différents problèmes lors d'une classification.

3.1. Problèmes liés aux classes

La classification supervisée suppose le choix de classes de références avant le traitement. Leurs caractéristiques sont fixées par les zones de calage qui leur sont associées et qui influencent directement le résultat de la classification.

Le problème est de trouver pour chaque classe une zone de calage représentative. Des paramètres tels que l'ensoleillement, le relief, ... font que la signature spectrale d'une même classe thématique varie suivant l'endroit où la zone de calage est choisie. On atténue l'influence de ces paramètres en prenant des pixels de la même classe thématique en plusieurs points dans l'image.

3.2. Problèmes liés au critère de ressemblance

Lors du processus de classification, les pixels seront répartis entre les classes d'après un critère de ressemblance.

"Que vaut-il mieux utiliser: une méthode basée sur un critère statistique ou un critère déterministe?"

Pour répondre à cette question, l'analyse par confrontation à la vérité-terrain reste la seule solution disponible.

La question devient alors:

"Dans quelle situation appliquer une méthode plutôt que l'autre?"

D'un point de vue purement théorique, l'utilisation d'une méthode déterministe simple ne nécessite la vérification d'aucune hypothèse et est donc en principe toujours justifiée.

Une méthode statistique impose que l'image est "poissonienne" ou "gaussienne".

Généralement, c'est un mélange des deux! En théorie, on ne pourrait donc pas appliquer ces méthodes, les hypothèses de base ne tenant pas.

En fait, les deux types de méthode fournissent des résultats plus ou moins satisfaisants et il n'existe pas de "loi" présidant au choix de l'une plutôt que l'autre *.

3.3. Temps de calcul

Un facteur qui peut entrer en ligne de compte dans le choix d'une méthode est le temps de calcul. Celui-ci augmente très rapidement avec la complexité de la procédure.

* se référer à Marc de Fays p°. 4 - 5

4. Exemples

4.1. Données

Les données dont nous disposons portent sur la région de WERIS (près de DURBUY en Belgique).

Cette région doit être répartie en sept catégories différentes (ville - résineux - feuillus - forêts mixtes - champs - prairies - friches).

Pour ce faire, nous disposons, pour chacun de ces groupes, de points de référence.

Le tableau ci-dessous donne le nombre de points de références disponibles dans chaque catégorie.

	CLASSES	POINTS DE REF.
1	ville	55
2	résineux	332
3	feuillus	1370
4	forêts mixtes	463
5	champs	1327
6	prairies	797
7	friches	840
	TOTAL	5184

A partir de ces points, nous devons classer tous ceux formant l'image. Il s'agit d'une image de 262 144 points (512 lignes * 512 colonnes). Pour chaque point (pixel), nous disposons de l'intensité lumineuse sur trois longueurs d'ondes différentes. L'intensité est une valeur entière s'échelonnant de 0 à 255.

...

4.2. Remarque

Les classifications réalisées pour les différentes méthodes ont été effectuées par V. GRANVILLE sur système VAX.

Nous avons classé les points selon la méthode du processus de Poisson non homogène, la méthode du processus de Poisson homogène et celle du maximum de vraisemblance.

4.3. Les cartes

Voici les cartes que nous obtenons:

la première carte est donnée par la classification par le processus de Poisson homogène;

la deuxième carte est la visualisation de la région par la classification par le processus de Poisson non homogène.

La signification des couleurs est la suivante:

	CLASSES	COULEURS
1	ville	rouge
2	résineux	vert foncé
3	feuillus	vert clair
4	forêts mixtes	vert
5	champs	orange
6	prairies	jaune
7	friches	violet
	inclassés	blanc

A titre de comparaison, la composition colorée ainsi que la carte obtenue par la méthode du maximum de vraisemblance sont fournies (attention, les couleurs ne sont pas les mêmes).

...

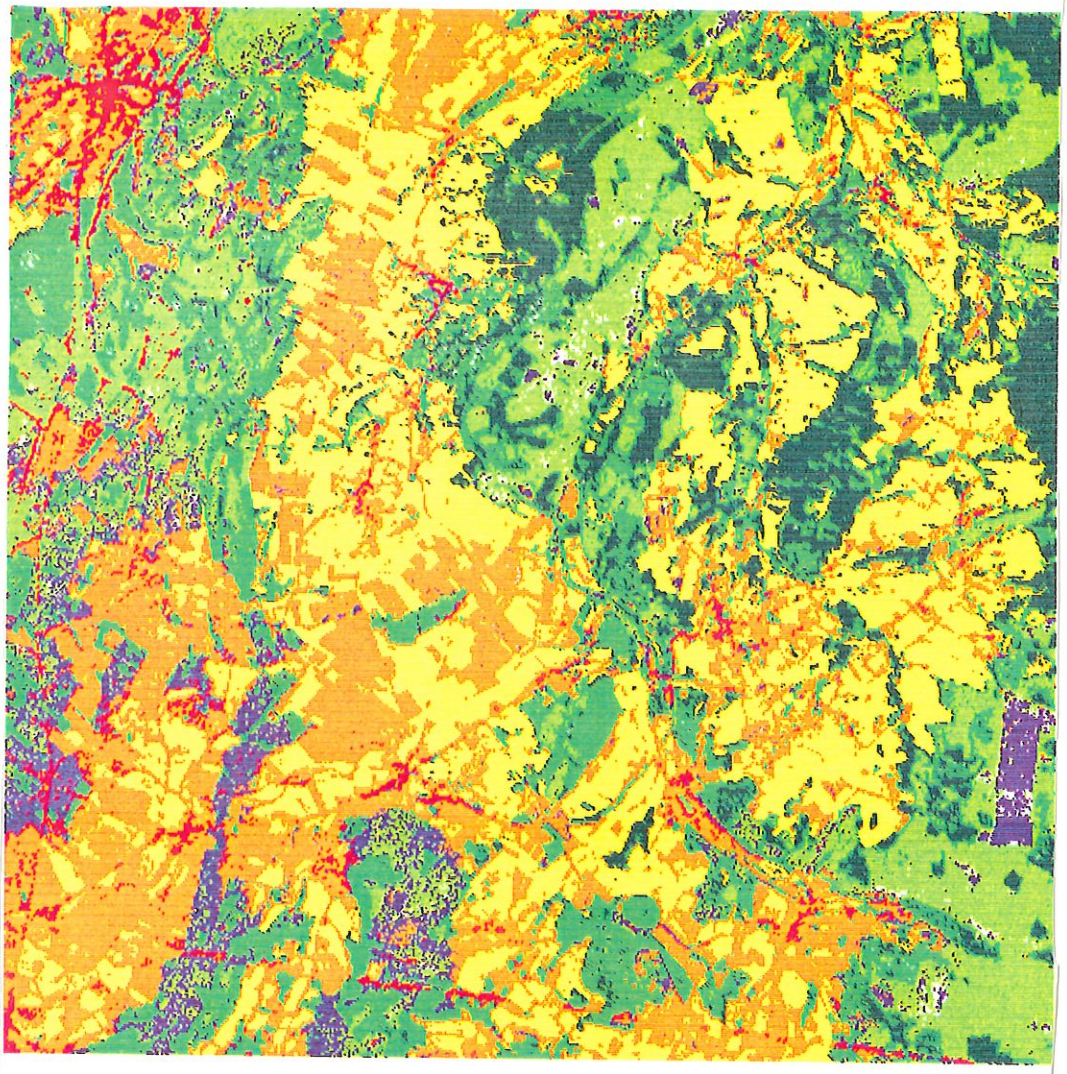


Fig. 1:

classification par le processus de Poisson homogène

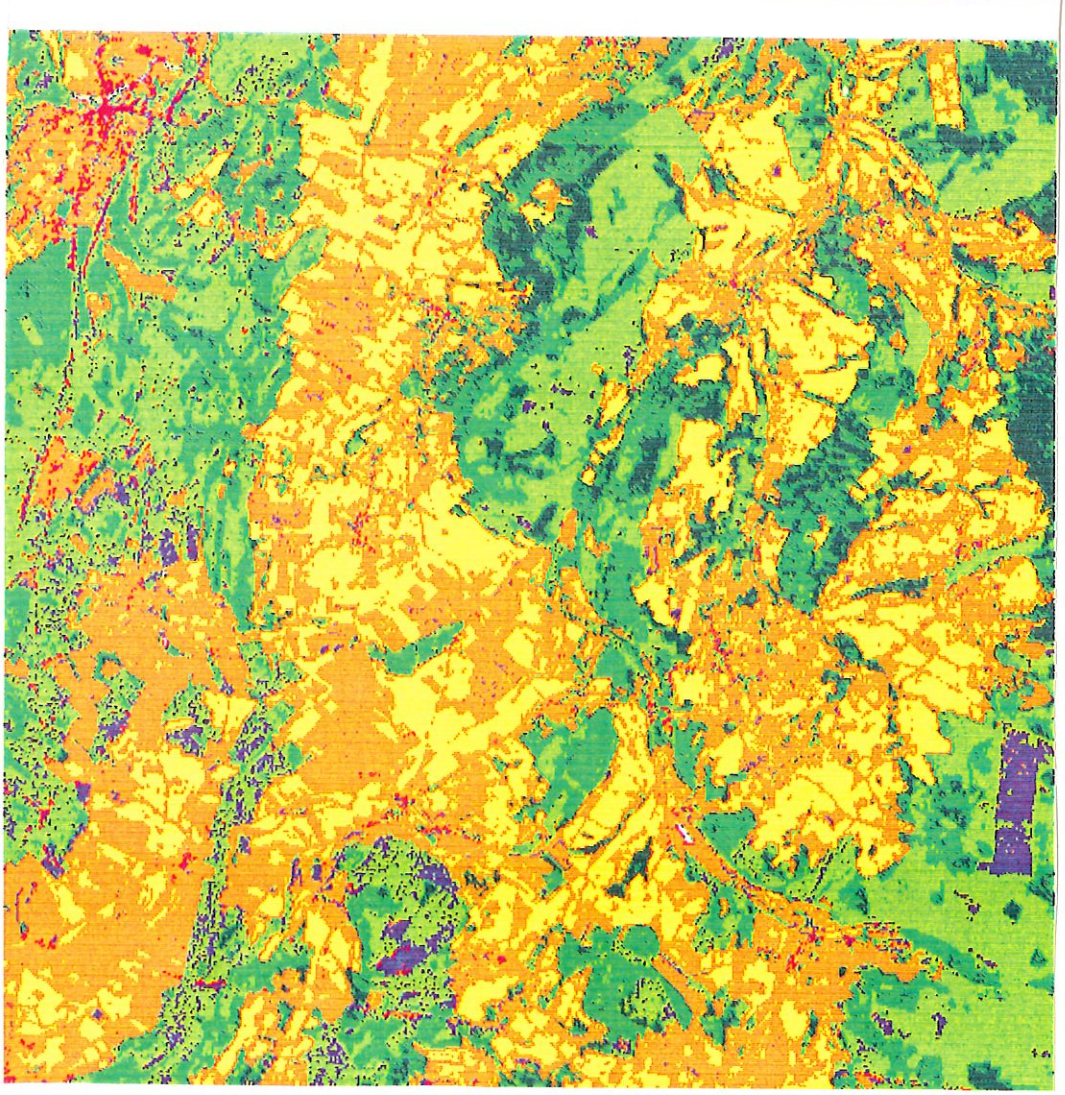


Fig. 2:

classification par le processus de Poisson non homogène

...

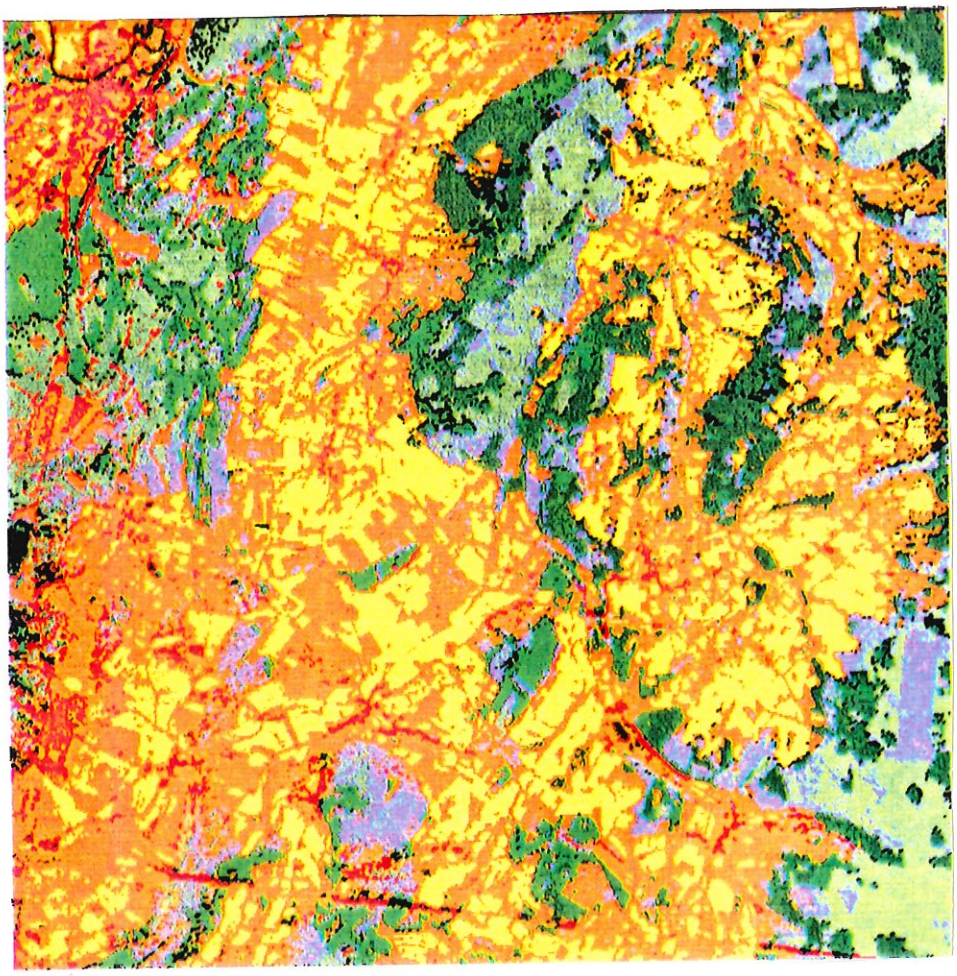


Fig. 3:

classification par maximum de vraisemblance

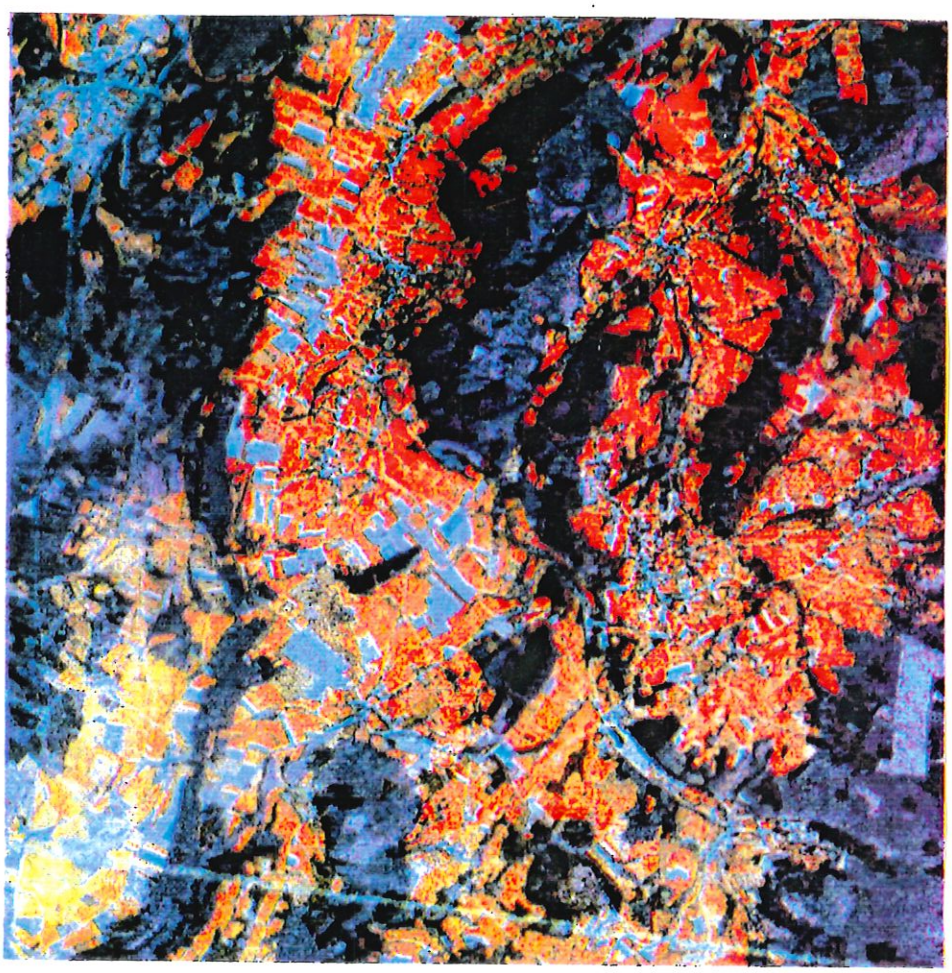


Fig. 4:

composition colorée

...

4.4. Classement des points

. On peut remarquer que des structures linéaires apparaissent dans le second cas. Ce sont probablement des routes pour lesquelles il n'y a pas d'ensemble d'entrainement.

. Dans le cas du processus de Poisson homogène, il y avait chevauchement des prairies et des champs. Maintenant, on voit mieux apparaître la proportion des champs.

. Dans la partie droite du dessous, les points non classés dans le cas du processus de Poisson homogène, sont classés par le processus de Poisson non homogène.

. Si nous comparons la carte obtenue par le processus de Poisson non homogène et celle du maximum de vraisemblance, nous pouvons constater:

- les extensions des forêts et la confusion pour le maximum de vraisemblance entre les forêts en vert et les friches dans la partie droite du dessous;

- l'incroyable différence entre les rapports de classifications entre les champs et les prairies (plus de 10 % de l'image) et la correction apportée par notre classification, où l'on peut suivre la classification champ par champ, presque pixel par pixel.

...

CONCLUSION

Du point de vue théorique, l'analyse discriminante suivant un processus de Poisson non homogène étend le domaine de recherche d'un modèle théorique pour le schéma entier d'une classification.

Ceci nous permet d'évoluer dans une direction très importante qui constituait un problème dans le cas du processus de poisson homogène.

Nous avons trouvé, sous certaines conditions décrites précédemment, un modèle théorique qui permet de classer les points appartenant à l'intersection de plusieurs domaines convexes. Ceci constitue un progrès très important pour la classification.

L'introduction de la notion d'intensité du processus fait intervenir des outils importants concernant l'estimation de densité. Notons que des travaux de recherche dirigés par RASSON améliorent les algorithmes, en utilisant des largeurs de bandes variables qui interviennent dans l'estimation des intensités.

Cette nouvelle méthode de classification est très prometteuse; nous pouvons suggérer à de futurs mémorisants de travailler dans cette direction afin de trouver des algorithmes plus performants ou de les améliorer.

REFERENCES ET BIBLIOGRAPHIES

(1) Baufays, P.

"Une nouvelle règle géométrique en analyse discriminante";

Thèse de doctorat, 1985;

Facultés Universitaires Notre Dame de la Paix,

Département de Mathématiques,

Namur.

(2) De Fays, M.

"Evaluation de la méthode de RASSON pour le traitement d'images par classification en télédétection";

Projet d'ingénieur, 1991;

Ecole Polytechnique Fédérale de Lausanne,

Suisse.

(3) Hardy A., Rasson J.P.

"Une nouvelle approche des problèmes de classification automatique";

Publication: Statistique et Analyse des données, 7, n° 2, p.p 41 - 56, (1982);

Facultés Universitaires Notre Dame de la Paix,

Département de Mathématiques,

Namur.

(4) Rasson, J.P.

"Estimation de domaines convexes dans le plan";

Publication: Statistique et Analyse des données, 1, p.p. 31 - 46 (1979);

Facultés Universitaires Notre Dame de la Paix,

Département de Mathématiques,

Namur.

...

[5] Rasson, J.P.

"Analyse discriminante pour le Processus de Poisson non homogène";

Rapport interne, 1992;

Facultés Universitaires Notre Dame de la Paix,

Département de Mathématiques,

Namur.

[6] Silverman, B.W.

"Density Estimation for Statistics and Data Analysis"

School of Mathematics, University of Bath, U.K.

Edité par Chapman and Hall (1986), Londres.

[7] Waerenburgh, L.

"Méthode exacte en analyse discriminante basée sur le critère de RASSON: application à la télédétection";

Mémoire, 1990;

Facultés Universitaires Notre Dame de la Paix,

Département de Mathématiques,

Namur.

TABLE DES MATIERES

Introduction 1

Formalisation du problème 2

Première partie: Processus de Poisson homogène 4

1. Introduction 4

2. Propriétés fondamentales du Processus de Poisson homogène 5

2.1. Hypothèses 5
2.2. Propriété 5
2.3. Remarque 5

3. Estimateur du maximum de vraisemblance d'un convexe D sur base d'une réalisation d'un Processus de Poisson homogène 7

3.1. Vraisemblance du modèle 7
3.2. Estimateur du convexe D 8

4. Analyse discriminante - Premier critère: domaines disjoints 9

4.1. Fonction de densité en un point du domaine 9
4.2. Règle de décision 11

4.2.1. Risque associé à chaque population 11
4.2.2. Risque bayesien 11
4.2.3. Conclusion 13

4.2.3.1. 1er Cas: le point à classer est à l'extérieur des enveloppes convexes 14

4.2.3.2. 2e Cas: le point à classer appartient à une seule enveloppe convexe 17

4.2.4. Exemple 18

5. Analyse discriminante - Second critère: domaines non disjoints 19

5.1. Premier cas: le point à classer n'appartient à aucune enveloppe convexe 19

5.2. Second cas: le point à classer appartient à une seule enveloppe convexe 19

5.3. Troisième cas: le point appartient à k enveloppes convexes 20

6. Conclusion 22

Deuxième partie: Processus de Poisson non homogène 23

1. Introduction 23

2. Estimateur du maximum de vraisemblance d'un convexe D sur base d'une réalisation d'un Processus de Poisson non homogène sur D 24

2.1. Vraisemblance du modèle 24
2.2. Estimateur du convexe D 25

3. Analyse discriminante - Premier critère: domaines disjoints 26

3.1. Fonction de densité en un point du domaine 26
3.2. Règle de décision 28

3.2.1. 1er Cas: le point à classer est à l'extérieur des enveloppes convexes 29

3.2.2. 2e Cas: le point à classer appartient à une seule enveloppe convexe 31

4. Analyse discriminante - Second critère: domaines non disjoints 32

5. Conclusion 33

Troisième partie: Méthodes d'estimation de densité 34

1. Introduction 34

2. Méthodes non paramétriques 35

2.1. Les histogrammes 35

2.1.1. Définition 35
2.1.2. Estimateur 35
2.1.3. Remarque 35
2.1.4. Utilité 35
2.1.5. Exemple 36

2.2. L'estimateur naïf 38

2.2.1. Construction 38
2.2.2. Estimateur 38
2.2.3. Interprétation 38
2.2.4. Exemple 39

2.3. L'estimateur des noyaux 41

2.3.1. Estimateur 41
2.3.2. Interprétation 41
2.3.3. Exemple 43

2.4. L'estimateur de densité du k^{ième}---- plus proche voisin 44

2.4.1. Estimateur 44
2.4.2. Interprétation 44
2.4.3. Exemple 45
2.4.4. Généralisation 45

2.5. Méthode du noyau variable 46

2.5.1. Estimateur 46
2.5.2. Interprétation 46
2.5.3. Exemple 47

2.6. Méthode du noyau adaptatif 48

2.6.1. Estimateur 48
2.6.2. Interprétation 48
2.6.3. Exemple 49

2.7. Estimateur de fonction de poids général 50

2.7.1. Estimateur 50
2.7.2. Utilité 50

3. Fonction noyau K(x) 51

4. Le noyau de l'estimateur naïf appliqué au cas discret 52

4.1. Formalisation du problème 52
4.2. Approche de BW SILVERMAN 53

4.2.1. Principe de départ 53
4.2.2. Choix du paramètre h 54
4.2.3. Bornes pour le paramètre h 54

4.2.3.1. Condition de recouvrement d'un point 54

4.2.3.2. Borne inférieure pour h 55

4.2.3.3. Borne supérieure pour h 58

4.2.3.4. Remarques 61

4.3. Problème important 66

Quatrième partie: Evaluation de la nouvelle méthode de classification pour le traitement d'images en télédétection 70

1. Introduction 70

2. Principe du traitement d'images par classification 71

2.1. Définition des classes 71
2.2. Critère 72
2.3. Formalisation du problème 73

3. Problématique du traitement d'image par classification 74

3.1. Problèmes liés aux classes 74
3.2. Problèmes liés au critère de ressemblance 75
3.3. Temps de calcul 75

4. Exemples 76

Conclusion 81

Références bibliographiques 82