



HAL
open science

Description statistique des données ordinales : analyse multidimensionnelle

Génard Drouet d'Aubigny

► **To cite this version:**

Génard Drouet d'Aubigny. Description statistique des données ordinales : analyse multidimensionnelle. Modélisation et simulation. Université Joseph-Fourier - Grenoble I, 1975. Français. NNT : . tel-00286221

HAL Id: tel-00286221

<https://theses.hal.science/tel-00286221>

Submitted on 9 Jun 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

N° D'ordre

T H E S E

présentée à

L'UNIVERSITE SCIENTIFIQUE ET MEDICALE DE GRENOBLE

pour obtenir

LE GRADE DE DOCTEUR DE 3EME CYCLE

"Mathématiques Appliquées"

par

Gérard D R O U E T D'A U B I G N Y

Description Statistique des Données Ordinales :

analyse multidimensionnelle

Thèse soutenue le 27 juin 1975, devant la Commission d'Examen.

M. BARRA	J.R.	Président
M. BENZAKEN	C.	Examineur
M. ESCOUFIER	Y.	Examineur
M. ROMIER	G.	Examineur

M. Michel SOUTIF : Président
M. Gabriel CAU : Vice-président

MEMBRES DU CORPS ENSEIGNANT DE L'U.S.M.G.

PROFESSEURS TITULAIRES

MM.	ANGLES D'AURIAC Paul	Mécanique des fluides
	ARNAUD Paul	Chimie
	AUBERT Guy	Physique
	AYANT Yves	Physique approfondie
Mme	BARBIER Marie-Jeanne	Electrochimie
MM.	BARBIER Jean-Claude	Physique expérimentale
	BARBIER Reynold	Géologie appliquée
	BARJON Robert	Physique nucléaire
	BARNOUD Fernand	Biosynthèse de la cellulose
	BARRA Jean-René	Statistiques
	BARRIE Joseph	Clinique chirurgicale
	BEAUDOING André	Clinique de Pédiatrie et Puériculture
	BERNARD Alain	Mathématiques Pures
Mme	BERTRANDIAS Françoise	Mathématiques Pures
MM.	BEZES Henri	Pathologie chirurgicale
	BLAMBERT Maurice	Mathématiques Pures
	BOLLIET Louis	Informatique (IUT B)
	BONNET Georges	Electrotechnique
	BONNET Jean-Louis	Clinique ophtalmologique
	BONNET-EYMARD Joseph	Pathologie médicale
	BOUCHERLE André	Chimie et Toxicologie
	BOUCHEZ Robert	Physique nucléaire
	BOUSSARD Jean-Claude	Mathématiques Appliquées
	BRAVARD Yves	Géographie
	CABANEL Guy	Clinique rhumatologique et hydrologie
	CALAS François	Anatomie
	CARLIER Georges	Biologie végétale
	CARRAZ Gilbert	Biologie animale et pharmacodynamie
	CAU Gabriel	Médecine légale et Toxicologie
	CAUQUIS Georges	Chimie organique
	CHABAUTY Claude	Mathématiques Pures
	CHARACHON Robert	Clinique Oto-Rhino-Laryngologique
	CHATEAU Robert	Thérapeutique (Neurologie)
	CHIBON Pierre	Biologie animale
	COEUR André	Pharmacie chimique et chimie analytique
	CONTAMIN Robert	Clinique gynécologique
	COUDERC Pierre	Anatomie Pathologique
	CRAYA Antoine	Mécanique
Mme	DEBELMAS Anne-Marie	Matière médicale
MM.	DEBELMAS Jacques	Géologie générale
	DEGRANGE Charles	Zoologie
	DELORMAS Pierre	Pneumo-Phthysiologie
	DEPORTES Charles	Chimie minérale
	DESRE Pierre	Métallurgie
	DESCAUX Georges	Physiologie animale
	DODU Jacques	Mécanique appliquée
	DOLIQUE Jean-Michel	Physique des plasmas
	DREYFUS Bernard	Thermodynamique
	DUCROS Pierre	Cristallographie
	DUGOIS Pierre	Clinique de Dermatologie et Syphillographie
	FAU René	Clinique neuro-psychiatrique

MM.	SAGNAIRE Didier	Chimie physique
	GALLISSOT François	Mathématiques Pures
	GALVANI Octave	Mathématiques Pures
	GASTINEL Noël	Analyse numérique
	GAVEND Michel	Pharmacologie
	GEINDRE Michel	Electroradiologie
	GERBER Robert	Mathématiques Pures
	GERMAIN Jean-Pierre	Mécanique
	GIRAUD Pierre	Géologie
	JANIN Bernard	Géographie
	KAHANE André	Physique générale
	KLEIN Joseph	Mathématiques Pures
	KOSZUL Jean-Louis	Mathématiques Pures
	KRAVTCHENKO Julien	Mécanique
	KUNTZMANN Jean	Mathématiques Appliquées
	LACAZE Albert	Thermodynamique
	LACHARME Jean	Biologie végétale
	LAJZEROWICZ Joseph	Physique
	LATREILLE René	Chirurgie générale
	LATURAZE Jean	Biochimie pharmaceutique
	LAURENT Pierre	Mathématiques Appliquées
	LEDRU Jean	Clinique médicale B
	LLIBOUTRY Louis	Géophysique
	LONGEQUEUE Jean-Pierre	Physique nucléaire
	LOUP Jean	Géographie
Mlle	LUTZ Elisabeth	Mathématiques Pures
	MALGRANGE Bernard	Mathématiques Pures
	MALINAS Yves	Clinique obstétricale
	MARTIN-NOEL Pierre	Sémiologie médicale
	MAZARE Yves	Clinique médicale A
	MICHEL Robert	Minéralogie et Petrographie
	MICOUD Max	Clinique maladies infectieuses
	MOURIQUAND Claude	Histologie
	MOUSSA André	Chimie nucléaire
	MULLER Jean Michel	Thérapeutique (néphrologie)
	NEEL Louis	Physique du Solide
	OZENDA Paul	Botanique
	PAYAN Jean-Jacques	Mathématiques Pures
	PEBAY-PEYROULA Jean-Claude	Physique
	RASSAT André	Chimie systématique
	RENARD Michel	Thermodynamique
	RINALDI Renaud	Physique
	DE ROUGEMONT Jacques	Neuro-chirurgie
	SEIGNEURIN Raymond	Microbiologie et Hygiène
	SENGEL Philippe	Zoologie
	SIBILLE Robert	Construction mécanique
	SOUTIF Michel	Physique générale
	TANCHE Maurice	Physiologie
	TRAYNARD Philippe	Chimie générale
	VAILLANT François	Zoologie
	VALENTIN Jacques	Physique Nucléaire
	VALQUOIS Bernard	Calcul électronique
Mme	VERAIN Alice	Pharmacie galénique
M.	VERAIN André	Physique
MM.	VEYRET Paul	Géographie
	VIGNAIS Pierre	Biochimie médicale
	YOCOCZ Jean	Physique nucléaire théorique

PROFESSEURS ASSOCIES

MM.	CHEEKE John	Thermodynamique
	COPPENS Philip	Physique
	CORCOS Gilles	Mécanique
	CRABBE Pierre	CERMO
	GILLESPIE John	I.S.N.
	POUYERREAU Daniel	Mathématiques Pures

PROFESSEURS SANS CHAIRE

Mlle	AGNIUS-DELORD Claudine	Physique pharmaceutique
	ALARY Josette	Chimie analytique
MM.	AMBROISE-THOMAS Pierre	Parasitologie
	BELORIZKY Elie	Physique
	BENZAKEN Claude	Mathématiques appliquées
	BERTRANDIAS Jean-Paul	Mathématiques appliquées
	BIAREZ Jean-Pierre	Mécanique
	BILLET Jean	Géographie
Mme	BONNIER Jane	Chimie générale
MM.	BOUCHET Yves	Anatomie
	BRUGEL Lucien	Energétique
	CONTE René	Physique
	DEPASSEL Roger	Mécanique des Fluides
	GAUTHIER Yves	Sciences biologiques
	GAUTRON René	Chimie
	GIDON Paul	Géologie et Minéralogie
	GLENAT René	Chimie organique
	GROULADE Joseph	Biochimie médicale
	HACQUES Gérard	Calcul numérique
	HOLLARD Daniel	Hématologie
	HUGONOT Robert	Hygiène et Méd. Préventive
	IDELMAN Simon	Physiologie animale
	JOLY Jean-René	Mathématiques pures
	JULLIEN Pierre	Mathématiques appliquées
Mme	KAHANE Josette	Physique
MM.	KUHN Gérard	Physique
	LOISEAUX Jean	Physique nucléaire
	LUU-DUC-Cuong	Chimie Organique
	MAYNARD Roger	Physique du solide
	PELMONT Jean	Biochimie
	PERRIAUX Jean-Jacques	Géologie et minéralogie
	PFISTER Jean-Claude	Physique du solide
Mlle	PIERY Yvette	Physiologie animale
MM.	RAYNAUD Hervé	M.I.A.G.
	REBECQ Jacques	Biologie (CUS)
	REVOL Michel	Urologie
	REYMOND Jean-Charles	Chirurgie générale
	RICHARD Lucien	Biologie végétale
Mme	RINAUDO Marguerite	Chimie macromoléculaire
MM.	ROBERT André	Chimie papetière
	SARRAZIN Roger	Anatomie et chirurgie
	SARROT-REYNAULD Jean	Géologie
	SIROT Louis	Chirurgie générale
Mme	SOUTIF Jeanne	Physique générale
MM.	VIALON Pierre	Géologie
	VAN CUTSEM Bernard	Mathématiques appliquées

MAITRES DE CONFERENCES ET MAITRES DE CONFERENCES AGREGES

MM.	AMBLARD Pierre	Dermatologie
	ARMAND Gilbert	Géographie
	ARMAND Yves	Chimie
	BARGE Michel	Neurochirurgie
	BEGUIN Claude	Chimie organique
Mme	BERIEL Héléne	Pharmacodynamique
M.	BOUCHARLAT Jacques	Psychiatrie adultes
Mme	BOUCHE Liane	Mathématiques (CUS)
MM.	BRODEAU François	Mathématiques (IUT B)
	BUISSON Roger	Physique
	BUTEL Jean	Orthopédie
	CHAMBAZ Edmond	Biochimie médicale
	CHAMPETIER Jean	Anatomie et organogénèse
	CHARDON Michel	Géographie
	CHERADAME Hervé	Chimie papetière

MM.	COHEN-ADDAD Jean-Pierre	Spectrométrie physique
	COLOMB Maurice	Biochimie médicale
	CORDONNIER Daniel	Néphrologie
	COULOMB Max	Radiologie
	CROUZET Guy	Radiologie
	CYROT Michel	Physique du solide
	DELOBEL Claude	M.I.A.G.
	DENIS Bernard	Cardiologie
	DOUCE Roland	Physiologie végétale
	DUSSAUD René	Mathématiques (CUS)
Mme	ETERRADOSSI Jacqueline	Physiologie
MM.	FAURE Jacques	Médecine légale
	FONTAINE Jean-Marc	Mathématiques Pures
	GAUTIER Robert	Chirurgie générale
	GENSAC Pierre	Botanique
	GIDON Maurice	Géologie
	GRIFFITHS Michaël	Mathématiques Appliquées
	GROS Yves	Physique (stag.)
	GUITTON Jacques	Chimie
	HICTER Pierre	Chimie
	IVANES Marcel	Electricité
	JALBERT Pierre	Histologie
	KOLODIE Lucien	Hématologie
	KRAKOWIAK Sacha	Mathématiques appliquées
Mme	LAJZEROWICZ Jeannine	Physique
MM.	LEROY Philippe	Mathématiques
	MACHE Régis	Physiologie végétale
	MAGNIN Robert	Hygiène et Médecine préventive
	MARECHAL Jean	Mécanique
	MARTIN-BOUYER Michel	Chimie (CUS)
	MICHOULIER Jean	Physique (I.U.T. "A")
Mme	MINIER Colette	Physique
MM.	NEGRE Robert	Mécanique
	NEMOZ Alain	Thermodynamique
	PARAMELLE Bernard	Pneumologie
	PECCOUD François	Analyse (IUT B)
	PEFFEN René	Métallurgie
	PERRET Jean	Neurologie
	PHELIP Xavier	Rhumatologie
	RACHAIL Michel	Médecine Interne
	RACINET Claude	Gynécologie et obstétrique
	RAMBAUD Pierre	Pédiatrie
Mme	RENAUDET Jacqueline	Bactériologie
MM.	ROBERT Jean Bernard	Chimie-Physique
	ROMIER Guy	Mathématiques (IUT B)
	SHOM Jean Claude	Chimie Générale
	STIEGLITZ Paul	Anesthésiologie
	STOEBNER Pierre	Anatomie pathologique
	VROUSOS Constantin	Radiologie

MAITRES DE CONFERENCES ASSOCIES

MM.	COLE Antony	Sciences nucléaires
	FORELL César	Mécanique
	MOORSANI Kishin	Physique

CHARGES DE FONCTIONS DE MAITRES DE CONFERENCES

MM.	BOST Michel	Pédiatrie
	CONTAMIN Charles	Chirurgie thoracique et cardio-vasculaire
	FAURE Gilbert	Urologie
	MALLION Jean Michel	Médecine du travail
	ROCHAT Jacques	Hygiène et hydrologie

Je tiens à remercier ici, tous ceux qui ont contribué à la réalisation de cette thèse, tant par leur enseignement, leurs conseils et critiques, que par leurs témoignages d'intérêt et d'amitié.

Je me suis "arraché à ma Bretagne natale" pour suivre les enseignements de Messieurs J.R. BARRA et N. GASTINEL, professeurs à l'Université Scientifique et Médicale de GRENOBLE. Aussi suis-je tout particulièrement sensible à l'honneur que me fait Monsieur BARRA en acceptant la présidence de ce Jury.

Trop de thèses sont passées par pur académisme, sans intérêt réel de l'auteur, pour que je ne remercie Monsieur G. ROMIER, professeur à l'Université des Sciences Sociales de GRENOBLE, de m'avoir proposé une si passionnante voie de recherche. Par son incitation constante à traiter des données réelles et son impressionnante faculté d'appréhension des problèmes, il m'a motivé à dépasser ma paresse naturelle et à mener à bien ce travail.

Je tiens à exprimer ma profonde reconnaissance à Monsieur C. BENZAKEN, professeur à l'Université Scientifique et Médicale de GRENOBLE, et à Monsieur Y. ESCOUFIER, Maître de Conférences à l'Université des Sciences et Techniques du Languedoc, pour l'intérêt qu'ils témoignent à mon travail, par leur participation à ce Jury.

Mes remerciements les plus sincères vont également à :

- Monsieur M. A. STEPHENS, professeur à l'Université Mac MASTER qui a dirigé mon stage de D.E.A. et dont les conseils amicaux m'ont convaincu d'entreprendre un travail de recherche,

- Monsieur G. MAUSER, professeur à l'Université de SAN DIEGO, qui m'a apporté un soutien patient et irremplaçable, par ses longues conversations et son aide bibliographique,

- Monsieur G. BAUMGARTNER, qui a mis à ma disposition les programmes disponibles à l'Institut d'Etudes Commerciales de GRENoble. Il m'a permis de mener à bien les nombreuses expérimentations informatiques qui ont jalonné et justifié ce travail.

La réalisation matérielle de ce document est due aux soins diligents du service de REPROGRAPHIE de l'I.U.T. de Statistiques de GRENoble, qui sous la direction attentive de Mademoiselle ANDREOLLE, vient parfaire la remarquable frappe de Madame GARCIN et de Mademoiselle BICAÏS. Je tiens à saluer ici, leur goût pour le travail bien fait.

TABLE DES MATIERES

INTRODUCTION

CHAPITRE I - LES MODELES DE DESCRIPTION DES DONNEES

- § 1..... Nature des données - Définitions et notations.
- § 2..... Typologie des échelles de mesure.
- § 3..... Modèles Interprétatifs.
- § 4..... Choix d'une métrique sur V .

CHAPITRE II - ELEMENTS DE REGRESSION ISOTONE

- § 1..... Régression isotone et projecteur dans L^2 .
- § 2..... Indice de régression isotone.
- § 3..... Algorithmes de régression isotone.
- § 4..... Le modèle linéaire ordinal.
- § 5..... Isotonie et préordonnance.

CHAPITRE III - ANALYSE DES DONNEES DE PROXIMITE AGREGES.

- § 1..... Le problème étudié en analyse des données de proximité globales.
- § 2..... L'approche heuristique de SHEPARD.
- § 3..... Formulation du problème en termes de régression isotone.
- § 4..... Les méthodes de Mc GEE et de de LEEUW.
- § 5..... Construction d'une configuration initiale.
- § 6..... L'analyse a-dimensionnelle des données de proximité.
- § 7..... L'analyse des données de proximité conditionnelles.

CHAPITRE IV - ANALYSE DES DONNEES DE PROXIMITE NON AGREGES.

- § 1..... Le problème étudié en analyse des données de proximité non-agrégées.
- § 2..... Résolution semi-métrique du problème.
- § 3..... Résolution purement ordinale du problème.

CHAPITRE V - ANALYSE DES DONNEES DE DOMINANCE

- § 1..... Le problème étudié en analyse des données de dominance.
- § 2..... Analyse interne : Méthodes de représentation conjointe.
- § 3..... Analyse interne : Méthodes de représentation factorielle.
- § 4..... Analyse externe des données de dominance.

BIBLIOGRAPHIE

I N T R O D U C T I O N

Le travail que nous présentons ici, concerne le domaine particulier de la statistique descriptive multidimensionnelle, consacré à la structuration et la description géométrique d'informations ordinales.

L'inadéquation des méthodes factorielles, au traitement de telles données, fut notée dès 1947 par THURSTONE ([16] p. XIII). Ce n'est cependant qu'en 1962 que SHEPARD [68,a] propose une méthode Heuristique d'analyse des DONNEES de PROXIMITE AGREGÉES. Le problème traité est le suivant :

Nous disposons de mesures de ressemblance (ou de dissemblance) entre éléments d'un ensemble fini d'objets I (Items, Stimuli, Hommes politiques, conditions expérimentales...), regroupées dans un tableau Δ . Ne retenant que la seule relation de préordre induite sur $I \times I$ par Δ , nous construisons une configuration de $\text{card}(I)$ points dans un espace vectoriel normé de dimension minimale qui soit invariante par toute transformation isotone des données et telle que les relations géométriques interpoints traduisent "au mieux" les relations significatives entre objets.

Depuis la formalisation du problème, en termes d'adéquation, par KRUSKAL [58,a] en 1964, de nombreuses approches ont été développées, tant en termes d'élasticité (Mc GEE 1966 [63,a]) , de permutations algébriques (GUTTMAN 1968 [52]) que de projection sur l'orthant positif (de LEEUW 1968 [33,c]).

La généralisation de ces méthodes à l'analyse des DONNEES de PROXIMITE non AGREGÉES, fut tentée dès 1968 (Mc GEE [63,b], KRUSKAL [58,d]). Une solution satisfaisante ne fut donnée qu'en 1970 par CARROLL [30,d]. Le problème, connu en psychométrie sous le vocable d'"Analyse des différences individuelles" est le suivant :

Pour chaque élément j d'un m -échantillon J issu d'une population donnée, nous disposons d'un tableau Δ^j de mesures de ressemblance entre éléments d'un ensemble fini d'objets, I . Nous construisons deux configurations dans un même espace vectoriel normé de dimension minimale :

i) une configuration de $\text{card}(I)$ points représentant, "au mieux", un compromis entre les préordres induits, sur $I \times I$, par les m tableaux Δ^j .

ii) une configuration de m points traduisant l'écart de chacun des préordres au compromis.

Parallèlement, l'extension des méthodes ordinales à l'analyse des DONNEES de DOMINANCE (CARROLL 1967 [15], ROSKAM 1968 [14], GLEASON 1968 [50]) permet de généraliser au cas multidimensionnel le modèle de DEPLOIEMENT de COOMBS [5].

Nous disposons alors, pour chaque élément j d'un m -échantillon J issu d'une population donnée, d'un classement (resp. de comparaisons par paires) des n éléments d'un ensemble d'objets I . Nous construisons une représentation simultanée des éléments de I et de J dans un espace vectoriel normé de dimension minimale, de telle sorte qu'elle soit invariante par toute transformation isotone des données et que les proximités géométriques entre points traduisent le degré de cohérence des classements (resp. comparaisons par paires) et permettent de reconstituer ceux-ci (resp. celles-ci).

L'analyse ordinale de ces trois types de données se distingue des méthodes usuelles de statistique descriptive multidimensionnelle ("multidimensional scaling") par l'affaiblissement de l'hypothèse de linéarité : une configuration est jugée représentative d'un ensemble d'objets si les distances interpoints sont fonction ISOTONE des indices de proximités correspondants. C'est pourquoi l'on ne retient que la représentation de plus faible dimension, qui reste invariante par toute transformation isotone des données.

Suivant la démarche de KRUSKAL [58,a], les différents auteurs construisent un indice d'adéquation d'un modèle ordinal aux données, et se proposent de l'optimiser. Cette démarche conduit à une succession de propositions algorithmiques sans fondement théorique commun. Un autre point

de vue sur ces méthodes, consiste à définir chacun des problèmes en termes de régression isotone. Cette démarche unificatrice est le point de départ de notre travail. Nous montrons ainsi les liens existant entre les trois types de problèmes, dégagés de leur contexte et langage psychométrique, et montrons de façon "naturelle" l'équivalence et les propriétés statistiques des indices d'adéquations proposés.

Cette démarche nous permet de montrer l'analogie des méthodes ordinales aux techniques usuelles de statistique descriptive multidimensionnelle et de montrer qu'il est possible d'analyser des données de proximité non agrégées par des méthodes purement ordinales.

Dans un premier chapitre, nous définissons trois critères de choix d'un (ou plusieurs) modèle(s) de description d'un tableau de données. Un premier critère est l'étude de la nature monaire ou binaire de la relation de ressemblance définie par les données. On retrouve ainsi la classification de COOMBS [5]. Le second critère reprend la distinction de STEVENS des différentes échelles de mesure des ressemblances et permet de définir le champ d'application des méthodes ordinales. Le troisième critère étudie le point de vue interprétatif qui fonde le modèle de description d'un tableau de données Δ . Nous reprenons ici la classification de SHEPARD [15]. Peu de choses sont nouvelles dans ce chapitre qui a pour but de convaincre le lecteur de l'importance méthodologique de la définition explicite d'un modèle descriptif.

Le deuxième chapitre étudie les éléments de régression isotone, nécessaires à notre exposé. Nous rappelons tout d'abord les résultats de BARLOW, BARTOLOMEW, BREMMER et BRUNK [2], sur la notion d'espérance conditionnelle par rapport à un σ -treillis. Seules ses propriétés de projecteur dans $L^2(\Omega, \mathcal{A}, \mathcal{P})$ sont données, lorsque Ω est muni d'une relation de préordre. Nous donnons alors deux algorithmes de calcul de cette espérance conditionnelle : l'algorithme de KRUSKAL [58, a] résoud le problème lorsque la relation de préordre est totale. dans le cas contraire nous exposons tout d'abord l'algorithme de GEBHARDT [48]. Cette approche pose en pratique des problèmes; aussi montrons nous qu'une formulation du problème en termes d'analyse multicritère, conduit à une solution utilisable en pratique. A cet effet, l'approche de de CANI [42], construisant la relation de préordre total, en termes de maximum de vraisemblance est opposée à la méthode de maximum de cohérence due à JACQUET LAGREZE [32].

Nous définissons ensuite un indice de corrélation isotone, par analogie avec le rapport de corrélation. Nous étudions alors en détail ses qualités d'optimalité. Les résultats de ce paragraphe sont nouveaux, et permettent de justifier les principaux indices d'adéquation proposés précédemment.

Enfin nous dégageons [33,b] la notion de modèle linéaire ordinal, permettant de donner un analogue ordinal aux méthodes de régression simple, multiple, à l'analyse factorielle discriminante et à l'analyse de variance. Une résolution algorithmique d'un tel modèle est exposée suivant les travaux de de LEEUW [33,b].

Le chapitre III étudie les données de proximité agrégées. Nous nous attachons, tout d'abord, à l'étude des théorèmes d'existence d'une solution forte au problème. Nous donnons une formulation et une démonstration nouvelles du théorème de HOLMAN [56], montrant la généralité de ce résultat et l'importance de l'hypothèse d'ultramétrie. Ce théorème nous conduit à opposer l'analyse des proximités à l'analyse hiérarchique et à montrer le besoin de tests statistiques, permettant d'opter pour l'un ou l'autre modèle en pratique. Nous exposons alors une démonstration simplifiée du théorème de GUTTMAN [52] donnant une condition suffisante d'existence d'une solution fortement isotone dans un espace euclidien de dimension moindre que $n - 2$ ($n = \text{card}(I)$).

L'approche heuristique de SHEPARD [68,a] est alors présentée, et interprétée en termes d'analyse de la dispersion, en reprenant les résultats de BENNETT [37]. Nous montrons que les formalisations ultérieures de KRUSKAL [58,a], GUTTMAN [52] et Mc GEE [63,a] s'attachent toutes à développer un aspect différent de l'approche de SHEPARD et opposons l'analyse de KRUSKAL fondée sur une régression isotone, à l'analyse de GUTTMAN. Une argumentation géométrique nous permet de montrer les qualités respectives de ces deux démarches.

La méthode de l'orthant positif due à de LEEUW [33,c] est présentée. Elle permet de choisir un critère d'ajustement autre que les moindres carrés et son interprétation statistique est donnée. Nous étudions ensuite la construction d'une initialisation des algorithmes itératifs exposés et montrons que notre démonstration du théorème de GUTTMAN permet un choix rationnel.

Le paragraphe suivant expose ce que nous appelons l'analyse a-dimensionnelle des données de proximité. Son but est de construire une description d'un tableau Δ par un tableau D de distances définies dans un espace de FINSLER. Nous montrons que notre approche formalise, en termes de régression isotone, sous contraintes, une proposition récente de SHEPARD et CUNNINGHAM [68,f].

Ce chapitre s'achève par la présentation des méthodes d'analyse des données de proximité conditionnelles. Le tableau de données Δ induit alors une famille $\{Q_i, i \in I\}$ de relations de préordre sur I, chaque élément i servant dans ce cas de référence pour construire Q_i . Nous formalisons ce modèle en termes de régression isotone.

Le problème posé en analyse des données de proximité non agrégées, est formalisé en termes de régression isotone au chapitre IV. Nous exposons les approches "métrique" et "semi-métrique" de CARROLL [43], issues de recherches en analyse factorielle trimodale. Aucune solution ordinale satisfaisante du problème n'a été donnée à ce jour. Notre formalisation montre que l'utilisation de techniques de régression isotone, permet de construire une représentation purement ordinale, par simple application du modèle linéaire ordinal. Nous proposons deux algorithmes résolutifs, dont l'un utilise une méthode due à SRINIVASAN et SHOCKER [72] pour construire un modèle aussi général que ceux de CARROLL présentés ci-dessus.

Le chapitre V est consacré à l'étude des données de Dominance. Nous formalisons le problème en termes de régression isotone et montrons que deux modèles représentatifs peuvent être construits par ces techniques. L'une est fondée sur la notion de point idéal, introduite par COCMB[5]. Elle construit une représentation conjointe des objets classés et des variables de classification, dans un même espace euclidien, par une analyse interne, id est limitée aux seules informations contenues dans les données de dominance. L'autre modèle d'analyse interne se propose de construire une représentation des objets et des variables en termes de dualité. Nous montrons qu'alors les techniques de régression isotone permettent de résoudre le problème grâce au modèle d'Analyse factorielle ordinale, étudié par SHEPARD-KRUSKAL [58,e] et ROSKAM [14].

Le paragraphe suivant montre que lorsqu'une représentation euclidienne des objets est connue, une analyse externe des données de préférence, permet

de construire une représentation conjointe des variables dans le même espace. CARROLL [30,c] propose une hiérarchie de quatre modèles d'analyse de telles données, lorsque celles-ci sont d'échelle intervalle. Lorsque les données sont d'échelle ordinale, nous opposons à l'analyse "semi-métrique" de CARROLL, une démarche purement ordinale, fondée sur les travaux de SRINIVASAN et SHOCKER [72].

En conclusion, c'est l'utilisation du langage de la régression isotone qui a permis de dégager les liens des différentes méthodes d'analyse ordinale entre elles et avec les méthodes usuelles de statistique descriptive multidimensionnelle, en termes d'analyse de la dispersion.

Elle met ainsi en évidence l'affaiblissement des hypothèses de linéarité en celles d'isotonie, qui se traduit en pratique par une moindre robustesse des méthodes ordinales. Le point commun à l'ensemble des méthodes descriptives multidimensionnelles étant de construire toujours une "solution", il est impératif de développer des méthodes statistiques de contrôle de la vraisemblance et de la "qualité" des représentations" obtenues.

CHAPITRE 1

LES MODELES DE DESCRIPTION DES DONNEES.

Le choix d'un modèle de description d'un tableau de données particulier nécessite une étude taxonomique des propriétés et de la structure des données elles-mêmes. Le chapitre en développant une telle analyse, nous permet de fixer le langage utilisé par la suite.

Le premier paragraphe étudie l'information contenue dans un tableau de données suivant le type de relation algébrique sous-jacente.

Au deuxième paragraphe nous utilisons la typologie de STEVENS [17] suivant le niveau de mesure, pour montrer la spécificité des méthodes ordinales parmi les techniques de statistique descriptive multidimensionnelle.

Un même tableau de données est susceptible de recevoir différentes interprétations. Nous distinguerons quatre modèles descriptifs fondamentaux, dans le paragraphe suivant. Nous montrons alors les limites de notre étude et donnons quelques références sur les modèles non étudiés dans la suite.

Le dernier paragraphe montre les difficultés de mise en oeuvre de techniques de représentation de données, en géométrie non euclidienne. Les chapitres suivants seront limités à l'étude de modèles euclidiens.

§ 1 - NATURE DES DONNEES - DEFINITIONS ET NOTATIONS.

On s'intéresse à deux ensembles finis de caractères définis sur une population Ω : I de cardinal n et J de cardinal m. En psychométrie, I représente en général un ensemble de stimuli (hommes politiques - items linguistiques - tests psychologiques - conditions expérimentales - couleurs - sons, etc...) et J un ensemble d'individus (personnes testées, interrogées, juges, etc...)

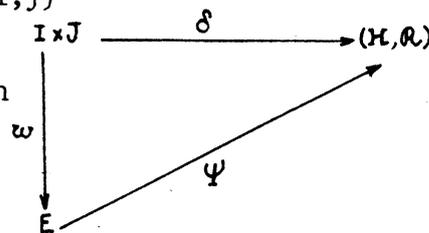
Dans des conditions expérimentales convenablement contrôlées $\omega(\omega \in \Omega)$, un élément (i,j) du produit cartésien I x J induit un effet :

$$e_{ij} = \omega(i,j)$$

élément de l'espace des états E, dont l'observation

$$\delta_{ij} = \Psi(e_{ij}) = \Psi \circ \omega(i,j)$$

élément de H, n'est contrôlée que par la relation p-aire \mathcal{R} , définie sur H. Les observations sont regroupées dans le tableau de donnée



$$\Delta(\omega) = \{\delta_{ij} ; (i,j) \in I \times J\}$$

dont l'étude nécessite une analyse de la relation \mathcal{R} . Nous retrouvons ainsi les classifications proposées par COOMBS [4] et ROSKAM [16].

I.1 - Définition

Un tableau de données $\Delta(\omega)$ est dit global si \mathcal{R} est une relation totale sur H. Il est dit conditionnel si \mathcal{R} est une relation partielle $\mathcal{R} = \cup \{ \mathcal{R}_i ; i \in I \}$ \mathcal{R}_i est une relation totale sur la section $H_i = \delta(\{i\} \times J)$

Le degré de la relation \mathcal{R} permet de distinguer quatre types de données globales :

* Lorsque \mathcal{R} est une relation monaire, l'observation δ_{ij} apporte une information sur le seul couple (i,j) de I x J. Nous avons alors la

I. 2 - Définition

Si \mathcal{R} est une relation totale monaire sur H , un tableau $\Delta(\omega)$ est une B-donnée (resp. une A-donnée) si $I = J$ (resp $I \neq J$) et alors $B = I \times I$ (resp. $A = I \times J$)

* De telles données ne sont pas abordées dans notre travail consacré au cas d'une relation binaire. induit alors une information sur les paires de dyades $((i_1, j_1), (i_2, j_2))$ éléments de $(I \times J)^2$ Nous posons alors la

I. 3 - Définition

Si \mathcal{R} est une relation totale binaire sur H , un tableau $\Delta(\omega)$ est une BxB-donnée (resp. une AxA-donnée) si $I = J$ (resp. $I \neq J$)

* Lorsque $\Delta(\omega)$ est une donnée conditionnelle, nous posons les définitions analogues :

I. 4 - Définition

Lorsque pour tout i dans I \mathcal{R}_i est une relation totale monaire sur H_i , un tableau $\Delta(\omega)$ est une B_i-donnée (resp. A_i-donnée) si $I = J$ (resp. $I \neq J$) et alors $B_i = \{i\} \times I$ (resp. $A_i : \{i\} \times J$)

I. 5 - Définition

Lorsque pour tout i dans I \mathcal{R}_i est une relation totale binaire sur H_i , un tableau $\Delta(\omega)$ est une $\bigcup (B_i \times B_i)$ donnée (resp. $\bigcup (A_i \times A_i)$ donnée)

Ces définitions permettent une première étude de la structure de données particulières - cependant même lorsque \mathcal{R} ou \mathcal{R}_i est une relation d'ordre, elles ne montrent pas la spécificité des méthodes ordinales. Les deux paragraphes suivantes répondent à ce but.

§ 2 - TYPOLOGIE DES ECHELLES DE MESURE

Le problème abordé par l'analyse multidimensionnelle peut être formulé comme suit : Etant donné un tableau $\Delta(\omega)$ de nombres réels représentant des mesures de ressemblance ou de dissemblance entre p caractères, construire une représentation Z de ces caractères dans un espace vectoriel normé V de telle sorte que les distances interpoints représentent "au mieux" les données suivant un critère d'ajustement. La définition de celui-ci est liée à l'échelle de mesure utilisée pour construire $\Delta(\omega)$. Nous distinguons de façon classique (c.f. [17]) six niveaux de richesse croissante :

2.1 - Echelle nominale

H est alors isomorphe à un sous-ensemble fini de \mathbb{N} et l'application Ψ est un codage des éléments de E . Seules sont admises les relations :

- d'équivalence entre états de même code (nous parlons d'exaequos)
- de non équivalence entre états de codes différents.

Les données catégorielles ou qualitatives sont mesurées suivant une telle échelle qui est invariante par permutation des codes.

2.2 - Echelle ordinale

H est isomorphe à un sous-ensemble de la droite réelle \mathbb{R} et \mathcal{R} est une relation de préordre notée \leq_H . Le code est alors hiérarchisé. δ est alors un indice de proximité caractérisant soit la ressemblance, soit la dissemblance de deux éléments de $I \times J$. Nous supposons dans toute la suite que δ exprime une notion de dissemblance.

I.6 - Définition

On appelle indice de dissimilarité, toute application δ définie sur $I \times J$ et à valeurs dans H qui, si $I = J$, vérifie pour tous (i, j) dans $I \times I$

$$D_1 : \delta_{ij} = \delta_{ji} \quad (\text{Symétrie})$$

$$D_2 : \delta_{ii} = \delta_{jj} \leq_H \delta_{ij} \quad \text{si } i \text{ et } j \text{ sont éléments de } \Omega .$$

La valeur numérique δ_{ij} n'est significative que du rang de l'état $w(i,j)$ dans la relation de préordre induite par Ψ sur E . Cette échelle est donc invariante par transformation isotone sur (H, \leq_H) .

I. 7 - Définition [1]

Soient (E, \mathcal{R}) et (E', \mathcal{R}') deux ensembles préordonnés. On appelle application isotone (ou isomorphisme de préordre) toute bijection M de E sur E' telle que :

$$(e_1 \mathcal{R} e_2) \Rightarrow M(e_1) \mathcal{R}' M(e_2) \quad \text{pour tous } e_1, e_2 \text{ dans } E.$$

La représentation Z d'un ensemble de p critères à partir de données d'échelle ordinale se fera donc suivant un critère d'isotonie :

$$d(Z(i), Z(j)) \leq d(Z(k), Z(l)) \quad \text{si } \delta_{ij} \leq_H \delta_{kl} \quad \text{pour tous } (i,j), (k,l) \text{ dans } I \times J$$

où d est la distance définie sur V . La mise en oeuvre de ce critère, depuis les travaux pionniers de SHEPARD [68,a] est étudiée par l'analyse multidimensionnelle ordinale.

2. 3 - Echelle hyperordinale

La notion de différence qualitative entre les valeurs de l'indice de dissimilarité est définie, ainsi qu'une relation de préordre sur ces valeurs.

Aucun critère d'adéquation particulier n'a été développé jusqu'ici pour de telles données, abordées en pratique comme des données ordinales.

2. 4 - Echelle d'intervalle

La notion de différence entre valeurs de l'indice de dissimilarité est définie numériquement. Elle est invariante par translation, aussi lorsque $I = J$ peut on imposer à δ d'être un indice de distance vérifiant la :

I. 8 - Définition

Une application T définie sur $I \times I$ à valeurs positives dans est un indice de distance si pour tous (i, j) de $I \times I$ elle vérifie :

$$T_1 : T_{ij} \geq 0$$

$$T_2 : (T_{ij} = 0) \Leftrightarrow i = j$$

$$T_3 : T_{ij} : T_{ji}$$

Une telle échelle est invariante par transformation affine et le "critère de la constante additive" développé par TORGERSON 17 permet d'analyser de telles données :

$$d(Z(i), Z(j)) = \delta_{ij} + c \quad \text{pour tout } i \text{ différent de } j.$$

2.5 - Echelle de proportion

Le quotient de deux valeurs est défini, conduisant à une échelle invariante par homothétie. δ est alors une distance et l'on peut analyser de telles données suivant un "critère d'égalité" :

$$d(Z(i), Z(j)) = \delta_{ij}$$

suivant la démarche de YOUNG et HOUSEHOLDER [17] .

2. 6 - Echelle de dénombrement

L'unité de mesure est alors fixée et seule l'identité laisse invariante une telle échelle.

Les méthodes usuelles de statistique descriptive multidimensionnelle étudient des données d'échelle d'intervalle. Les échelles ordinales et hyperordinales sont abordées par les méthodes ordinales exposées dans ce travail. Cependant les modèles d'analyse de telles données diffèrent suivant l'interprétation donnée au tableau $\Delta(\omega)$.

§ 3 - MODELES INTERPRETATIFS.

La description statistique de données particulières nécessite la définition d'un ou de plusieurs modèles interprétatifs. Ceux-ci permettent d'établir une correspondance simple entre relations inter-caractères exprimées par les données et relations inter-points dans l'espace représentatif. Nous distinguons quatre grands modèles, suivant en cela COOMBS [5] et SHEPARD [15].

Nous précisons les notations du paragraphe précédent en distinguant dans l'espace métrique (V, d) la représentation $X = \{x(i), i \in I\}$, de celle de $J : Y = \{y(j), j \in J\}$. Nous notons alors $d_{ij} = d(x(i), y(j))$ pour tous (i, j) dans $I \times J$.

3. 1 - Modèle de Proximité

Interpréter un tableau $\Delta(\omega)$ en termes de proximité, c'est considérer que la distance d_{ij} mesure dans \mathcal{R} le degré de dissemblance δ_{ij} des caractères i et j . Ceci suppose l'existence d'une relation isotone :

$$D(\omega) = M(\Delta(\omega)) \quad \text{i.e.} \quad d_{ij} = M(\delta_{ij}) \quad (\forall (i, j) ; (i, j) \in I \times J)$$

qu'il convient de mettre en évidence et de construire. Tel est le but fixé à l'analyse des proximités par SHEPARD [68, a].

Deux types de tableaux de données sont directement interprétables en ces termes. Lorsque l'on étudie des BxB-données, $\Delta(\omega)$ est un tableau carré d'ordre n (où $n = \text{card}(I)$) tel que δ_{ij} mesure la similarité, le degré de substitution, l'affinité, la confusion, l'association, la corrélation ou encore l'interaction entre deux stimuli.

De même les U(BixBi) données sont étudiées suivant ce modèle et le tableau $\Delta(\omega)$ se présente comme dans le cas précédent. La définition de \mathcal{Q} dans ce cas nécessite un affinement de nos notations : la relation isotone sous-jacente s'écrit alors :

$$d_{ij} = M_i(\delta_{ij}) \quad (\forall (ij), (i,j) \in \{i\} \times I)$$

Ce type de données est souvent rencontré en pratique, où il correspond à des méthodes très fiables de recueil des données, telles la méthode des tryades [1] ou celle des points d'ancrage [10].

La notion de "point-idéal" introduite par COOMBS [5] permet d'interpréter les données de préférence suivant ce modèle. Le tableau $\Delta(\omega)$ a alors n lignes et m colonnes et constitue une $U(A_j \times A_j)$ -donnée. Le terme général δ_{ij} mesure la proximité d'un caractère i au caractère j . Ce peut être le degré de préférence de l'individu j pour l'objet i ou la fréquence de fréquentation du magasin i par l'acheteur j etc...

Cette interprétation des $U(A_j \times A_j)$ -données conduit aux méthodes de déploiement développées dans le cas unidimensionnel par COOMBS [5] dont nous étudions les généralisations au cas multidimensionnel.

3. 2 - Le modèle de Dominance.

Interpréter un tableau $\Delta(\omega)$ en termes de dominance, c'est ordonner les éléments de I selon leur degré de possession d'une propriété j ($j \in J$). Deux types de données sont interprétables en ces termes.

Les relations de Tournois ou de comparaison par paire conduisent à des A_j -données. Le terme général δ_{il} du tableau carré d'ordre n $\Delta(\omega)$ mesure le degré de dominance de i par l . Ce sera par exemple le nombre de

fois que 1 a battu i ou bien le score de i face à 1 dans un tournoi de tennis. j représente la force d'un joueur au tennis et correspond à l'ensemble du tableau $\Delta(\omega)$.

Lorsque l'on est assuré de la transitivité de la relation de dominance ainsi définie, une telle approche conduit à une représentation unidimensionnelle (c.f KLEMMER et SCHRIMPTON [15]) dans le cas contraire il est nécessaire de construire une représentation multidimensionnelle, suivant la démarche de HAYASHI [55] par exemple. Cependant une démarche algébrique conduisant à une description par graphe nous semble plus satisfaisante. De telles données seront ignorées dans la suite.

Lorsque nous disposons de plusieurs tableaux de comparaisons par paire, ou de plusieurs classements, l'absence de consensus strict conduit à l'utilisation de techniques de description multidimensionnelle de $\mathcal{V}(A_j \times A_j)$ -données, dont les données de préférences étudiées ci-dessus sont un exemple.

Dans la suite nous réservons le terme de modèle de dominance à l'étude de $\mathcal{V}(A_j \times A_j)$ données.

3. 3 - Modèle de Profils.

Au terme ambigu de "données multivariées" nous préférons celui de "donnée de profil" et donnons la

I. 9 - Définition

On appelle profil de l'élément $i \in I$ (resp. $j \in J$) le tableau de description de i (resp. j) par les éléments de J (resp. I) et notons :

$$\delta_i = \{\delta_{ij}, j \in J\} \quad \delta_j = \{\delta_{ij}; i \in I\}$$

Interpréter un tableau () en termes de profils c'est admettre l'existence d'une relation fonctionnelle simple entre les coordonnées du vecteur $x(i)$ (resp. $y(j)$) représentant l'élément $i \in I$ (resp. $j \in J$) dans (V, d) et le profil δ_i . (resp. δ_j).

Cette relation fonctionnelle peut être linéaire comme en analyse factorielle classique ou en analyse en composantes principales ([3], [16], [17]) ou polynomiale comme en analyse factorielle non linéaire [62] où GNANADESIKAN et WILK in KRISHAIAH 1969). Plus généralement elle peut être isotone ([33,b], [53], [58, e], [68, e]) ou "continue" au sens de CARROLL ([68, e]).

Ce type de modèle est sans doute le plus usité puisqu'il est le fondement de toutes les méthodes de description statistique multidimensionnelle de données d'intervalle.

3. 4 - Modèle des mesures conjointes

Interpréter un tableau $\Delta(\omega)$ en termes de mesures conjointes c'est considérer que chaque élément δ_{ij} est une fonction simple de deux coordonnées correspondant respectivement à une représentation unidimensionnelle disjointe de I et de J.

Cette fonction peut être additive comme en analyse de la variance, multiplicative ou plus généralement polynomiale (c.f. LINGOES et YOUNG in [15] CARROLL in [11]), ou encore isotone ([33,b], [58,b]).

Un tel modèle, surtout développé dans le cadre de données d'échelle nominale sera pas étudié dans la suite.

L'exemple de l'analyse des données de préférence montre qu'un même tableau de données peut être interprété suivant plusieurs modèles. De plus à l'exception du modèle des mesures conjointes, les différents modèles donnent des informations complémentaires sur la structure de données et seule l'interprétation des résultats permet de dire a posteriori quel est le modèle le mieux adapté à la description des données, s'il en existe un.

§ 4 - CHOIX D'UNE METRIQUE SUR V.

La description de données ordinales, s'est développée suivant deux voies, à partir de 1962. La première fondée sur les travaux de FLAMENT [46] aborde le problème en termes de théorie des graphes et ne sera pas abordée ici.

Nous étudions l'approche géométrique du problème, fondée sur les travaux de SHEPARD [68,a]. Les données définissent un espace semi-métrique (la semi-métrique est définie par tout indice de distance isotone à δ) qu'il convient de plonger dans un espace métrique V.

Lorsque l'on impose de plus à V d'être ultramétrique, le problème est résolu par l'analyse hiérarchique [3]. Dans le cas contraire V est un espace de FINSLER dans lequel le problème n'a pas reçu de réponse dans sa généralité (c.f; [4]).

Cependant les travaux axiomatiques de BEALS, KRANTZ et TVERSKY [36] [76 a et b]) définissent les propriétés qualitatives minimales que doivent vérifier les données de proximité pour être représentées dans un espace de MINKOWSKI où la distance d est définie par :

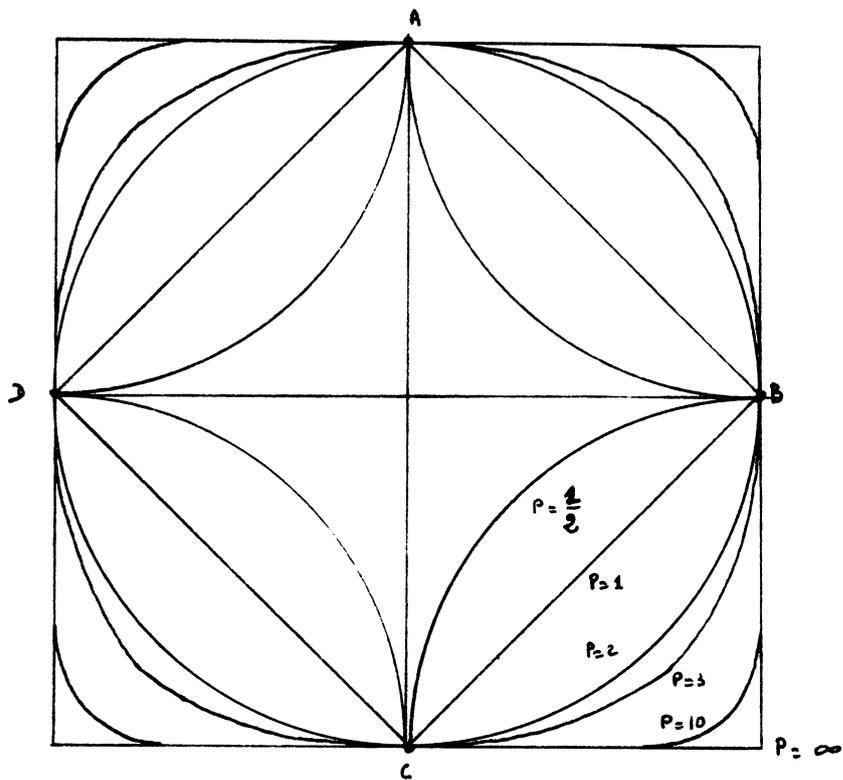
$$d(x(i), x(j)) = \left[\sum_{k=1}^r |x_k(i) - x_k(j)|^p \right]^{1/q} \quad p, q \in \overline{\mathbb{R}}_+$$

où l'on a supposé V r dimensionnel. On note L_{pq} une telle métrique.

Les espaces sont trop généraux et bien que de LEEUW [33-c] ait montré l'intérêt statistique d'une étude menée dans V muni d'une métrique L_{pq} avec $0 < p \leq 1$ et $q = 1$, les psychométriciens n'étudient le problème que dans les espaces de HÖLDER où $p = q$ - ($L_{pp} = L_p$).

Les jauges associées à ces métriques (Fig.1) prennent alors une grande importance puisqu'elles conditionnent la forme des régions d'isotonie et donc la lecture des représentations graphiques construites en cours d'étude.

Leur étude permet de tracer les hyperplans médiateurs associés à deux points A et B dont la complexité interdit toute lecture naïve d'une représentation graphique lorsque la métrique de HÖLDER n'est pas euclidienne (Fig.2).



BOULES UNITES ASSOCIÉS
AUX DISTANCES DE HÖLDER

Figure 1

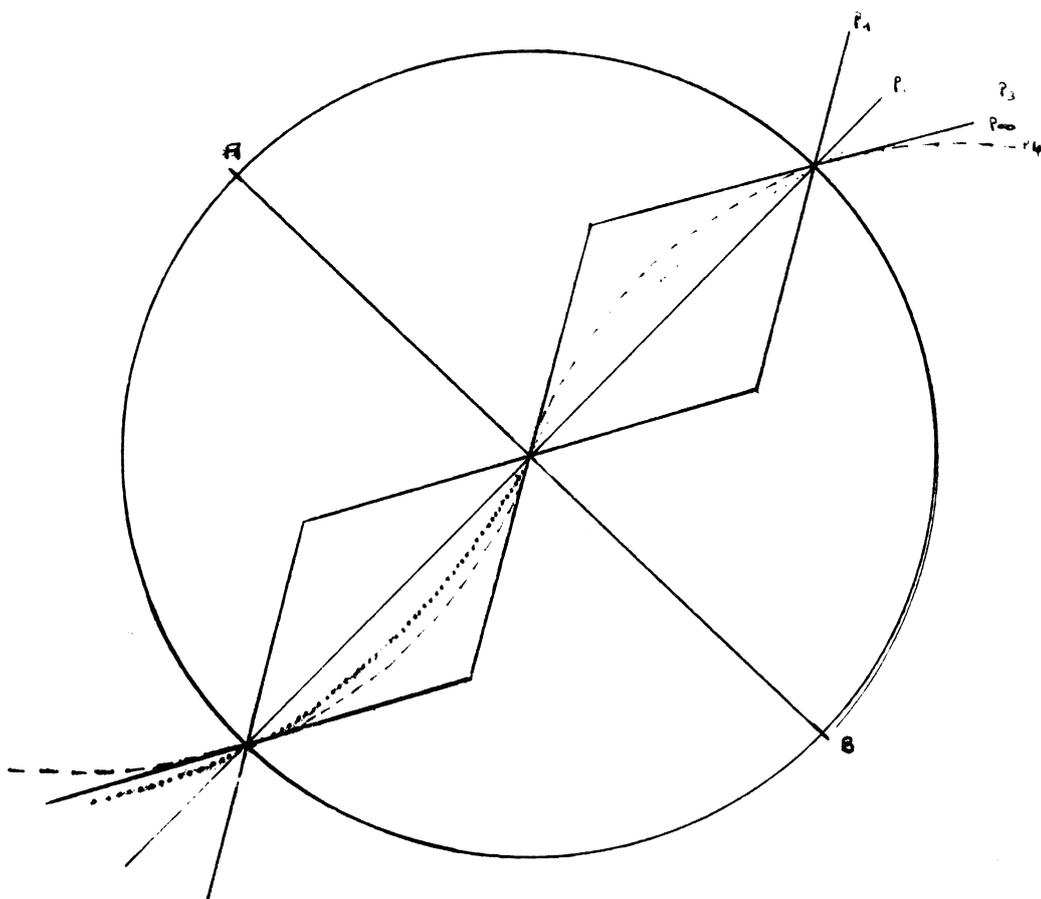
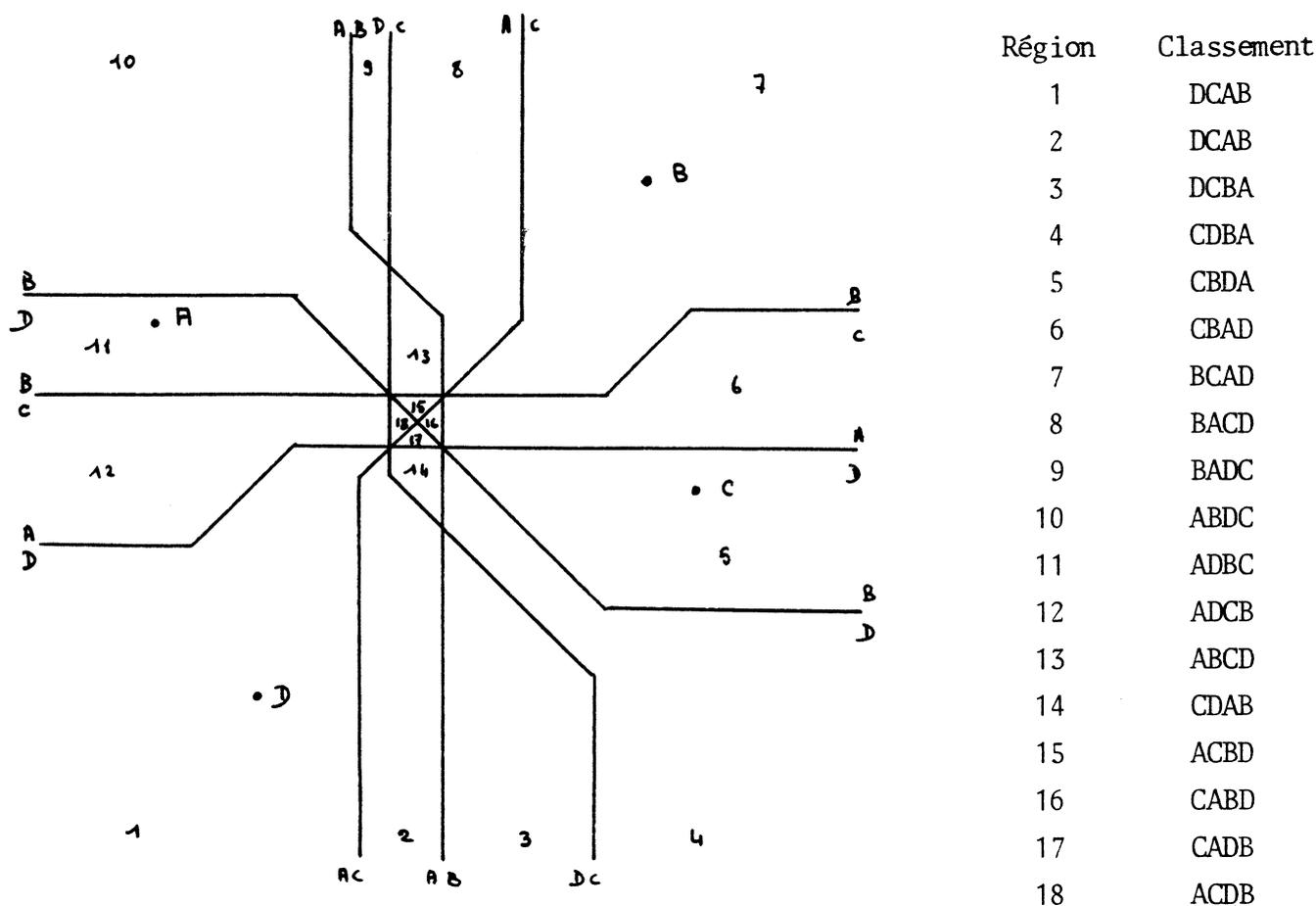


Figure 2

Hyperplans médiateurs du segment AB.

Nous reproduisons ci-dessous une configuration de quatre points en métrique L_1 . Nous avons tracé les hyperplans médiateurs de chaque segment afin de montrer la complexité de lecture d'une telle carte.



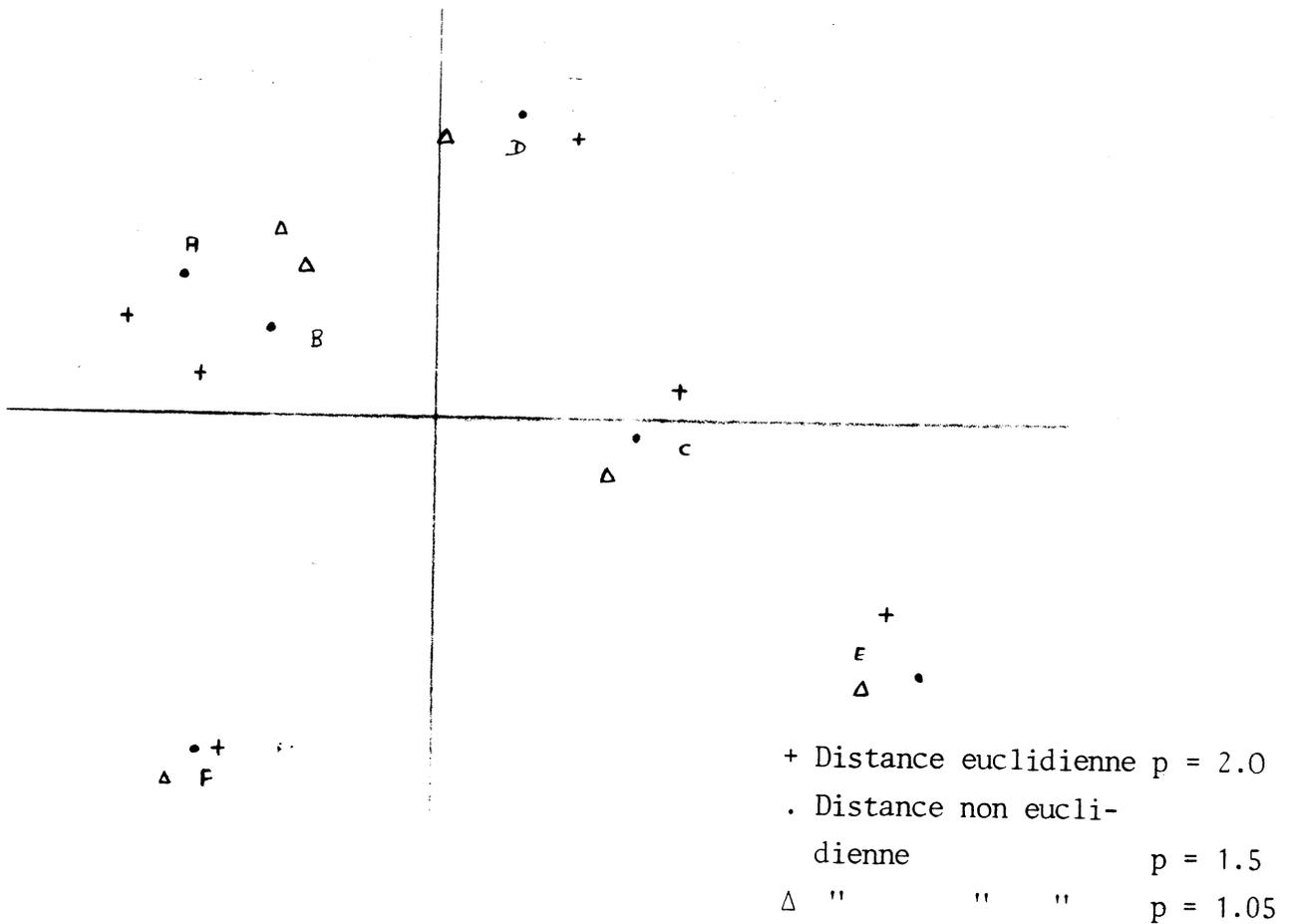
Le choix d'une métrique l_p est parfois motivé par une information a priori. Tel est le cas d'études psychométriques en distance L_1 ou L_2 (c.f. [68,c] pour une bibliographie).

Lorsqu'il n'en est pas ainsi, ceci est un paramètre supplémentaire du modèle, qu'il convient d'estimer. Lorsque la dimension r de V est fixée et un indice d'adéquation du modèle aux données établi, la démarche usuelle consiste à construire des représentations des données pour diverses valeurs de p et retenir celle qui minimise l'indice d'adéquation.

Les risques de mauvaise estimation préalable de r , la sensibilité des méthodes ordinales aux minima locaux, condamnent en fait une telle

méthode. La solution consistant à faire pour chaque valeur de p plusieurs tentatives devient si coûteuse qu'elle ne peut être qu'exceptionnellement employée.

L'influence du choix de p sur la structure de la représentation obtenue n'a été que peu étudiée et ce par simulation ([68-a], [69]). Les recherches montrent que la configuration est peu modifiée comme en témoigne la figure ci-dessous, reproduite comme les précédentes de ROSKAM [14].



Cependant les algorithmes existant actuellement ont une propension accrue à construire des configurations dégénérées lorsque p est différent de deux.

Nous avons par suite limité notre propos à l'étude du cas où V est un espace euclidien (i.e. $l = 2$).

CHAPITRE 2

ELEMENTS DE REGRESSION ISOTONE

La description multidimensionnelle des données ordinales fut rendue possible par SHEPARD [68,a]. Son approche heuristique a reçu de nombreuses formalisations statistiques en termes d'adéquation d'un critère d'isotonie aux données ([33-c], [52], [58-a], [63], [80,a]). Cependant l'utilisation de techniques de régression isotone pour minimiser la fonction d'écart ("Badness of-fit function") n'a, semble-t-il jamais retenu l'attention des auteurs. Notre travail montre le rôle central et unificateur de la régression isotone en analyse des données ordinales et permet de fonder statistiquement les choix de fonctions d'écarts proposés antérieurement. Dans ce but, nous exposons dans ce chapitre les outils de régression isotone nécessaires à notre exposé.

Le premier paragraphe, reprenant les résultats de BARLOW, BARTOLOMEW, BREMMER et BRUNK [2] limite l'étude de la régression isotone à ses aspects géométriques. En particulier le problème de l'existence de l'espérance conditionnelle par rapport à un σ -treillis, en termes de dérivée de Radon-Nikodym, n'est pas abordé. Nous renvoyons le lecteur intéressé par un exposé complet à [2].

Nous étudions dans le second paragraphe la définition et les propriétés élémentaires d'un indice de régression isotone qui joue dans ce cadre un rôle analogue à celui du rapport de corrélation dans le cas usuel. La généralisation au cas isotone des travaux développés en cartographie par GILBERT [49], nous permettent de donner quelques résultats d'optimalité de l'indice retenu.

La mise en oeuvre algorithmique de méthodes de régression isotone est étudiée au paragraphe trois. Les algorithmes exposés sont différenciés suivant les problèmes spécifiques rencontrés lorsque le σ -treillis est engendré par une ou par plusieurs variables aléatoires.

Le paragraphe suivant étudie l'extension au cas ordinal du modèle linéaire : On étudie alors la description d'une variable aléatoire ordinale par une famille finie de variables quantitatives. Nous définissons ainsi les équivalents ordinaux des méthodes de régression, d'analyse factorielle discriminante et de l'analyse de la variance.

Le dernier paragraphe, consacré à l'étude des relations d'isotonie entre préordonnances, introduit la notion de signature d'une préordonnance due à GUTTMAN [52] que nous utilisons dans les chapitres suivants. Trois types d'isotonie sont différenciés suivant le traitement souhaité des exaequos.

§ 1 - ELEMENTS DE REGRESSION ISOTONE.

On étudie une variable aléatoire Y définie sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathcal{P})$ et à valeurs dans $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ où $\mathcal{B}_{\mathbb{R}}$ est la tribu borelienne de \mathbb{R} . On suppose Y de carré intégrable et on note : $Y \in L^2(\mathcal{A})$. On suppose définie une variable aléatoire X définie sur $(\Omega, \mathcal{A}, \mathcal{P})$ à valeurs dans (E, \mathcal{E}) où E est un ensemble préordonné par la relation \mathcal{R} .

II. 0 - Définition

On appelle variable ordinale, toute variable aléatoire définie sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathcal{P})$ à valeurs dans un ensemble préordonné (E, \mathcal{R}) .

Le but poursuivi est de décrire Y par une application mesurable de (Ω, \mathcal{A}) dans $(\overline{\mathbb{R}}, \mathcal{B}_{\overline{\mathbb{R}}})$ qui factorise à travers X . Si l'on note \mathcal{H} l'ensemble des fonctions isotones m de (E, \mathcal{R}) dans $\overline{\mathbb{R}}$ muni de la relation d'ordre usuel, telles que $f = m \circ X$ appartienne à $L^2(\mathcal{A})$. On se propose de déterminer l'élément \hat{Y} dans \mathcal{H} qui minimise :

$$\int_{\Omega} (Y(\omega) - f(\omega))^2 P(d\omega)$$

En effet, si nous définissons sur l'espace probabilisable $(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}})$ la probabilité P_Y image de P par Y , et supposons donnée une transition de probabilité $\nu(\cdot, \cdot)$ sur $\mathcal{B}_{\bar{\mathbb{R}}} \times \Omega$, nous sommes assurés de l'existence d'une mesure positive λ définie sur $(\Omega \times \bar{\mathbb{R}}, \mathcal{G} \otimes \mathcal{B}_{\bar{\mathbb{R}}})$ (NEVEU prop. III Dunod 1970).

$$\lambda(A \times B) = \int_A \nu(B, \omega) P(d\omega) \quad \forall A \times B, A \times B \in \mathcal{G} \otimes \mathcal{B}_{\bar{\mathbb{R}}}$$

Si nous notons π_Ω (resp. $\pi_{\bar{\mathbb{R}}}$) l'application coordonnée de $\Omega \times \bar{\mathbb{R}}$ sur Ω (resp. $\bar{\mathbb{R}}$) :

$$\nu(B, \omega) = P_{\pi_{\bar{\mathbb{R}}}} \{ \pi_{\bar{\mathbb{R}}} \in B \mid \pi_\Omega = \omega \}$$

car $\nu(B, \omega)$ est une probabilité de transition, par suite :

$$\begin{aligned} \textcircled{*} \quad \int_{\Omega \times \bar{\mathbb{R}}} (y - mX(\omega))^2 \lambda(dy, \omega) &= \int_{\Omega} P(d\omega) \int_{\bar{\mathbb{R}}} (y - mX(\omega))^2 \nu(dy, \omega) \\ &= \int_{\Omega} E \{ (\pi_{\bar{\mathbb{R}}} - mX(\omega))^2 \mid \pi_\Omega = \omega \} P(d\omega) \end{aligned}$$

Si l'on note $Y(\omega) = E(\pi_{\bar{\mathbb{R}}} \mid \pi_\Omega = \omega)$ pour tout ω dans Ω

On remarque que :

$$E \{ (\pi_{\bar{\mathbb{R}}} - mX(\omega))^2 \mid \pi_\Omega = \omega \} = E \{ (\pi_{\bar{\mathbb{R}}} - Y(\omega))^2 \mid \pi_\Omega = \omega \} + (Y(\omega) - mX(\omega))^2$$

et donc maximiser $\textcircled{*}$ équivaut à maximiser :

$$\int_{\Omega} (Y(\omega) - mX(\omega))^2 P(d\omega)$$

Ce problème fut étudié en détail par BARLOW, BARTOLOMEW, BREMMER et BRUNK ([2]) et les résultats exposés dans ce paragraphe sont démontrés dans ([2] ch. VII pp. 307-358).

II. 1 - Définition

On appelle section finissante sur E tout élément \mathcal{U} de \mathcal{E} dont l'indicatrice est une fonction isotone de (E, \mathcal{R}) dans $([0,1], \leq)$ où " \leq " est la relation d'ordre usuelle sur $[0,1]$. On note \mathcal{U}_E la classe des sections finissantes sur E .

Remarque :

Si l'on se reporte à la définition I.7 on note que \mathcal{U} est une section finissante si et seulement si :

$$(e_1 \mathcal{R} e_2) \Rightarrow (\bigcap_{\mathcal{U}}(e_1) \leq \bigcap_{\mathcal{U}}(e_2)) \quad \text{pour tous } e_1, e_2 \text{ dans } E.$$

ce qui équivaut à :

$$((e_1 \mathcal{R} e_2) \text{ et } (e_1 \in \mathcal{U})) \Rightarrow (e_2 \in \mathcal{U})$$

justifiant ainsi le terme de section finissante.

Si l'on n'imposait pas l'inclusion de \mathcal{U}_E dans \mathcal{E} , il est clair que \mathcal{U}_E serait un treillis complet. Cependant \mathcal{E} n'est close que par union et intersection dénombrables, aussi \mathcal{U}_E n'est-il qu'un σ -treillis si l'on impose cette condition de mesurabilité.

II. 2 - Définition

Une application $f : (E, \mathcal{E}) \rightarrow (\overline{\mathbb{R}}, \mathcal{B}_{\overline{\mathbb{R}}})$, où $\mathcal{B}_{\overline{\mathbb{R}}}$ est la tribu borelienne de la droite achevée munie de sa topologie usuelle, est dite \mathcal{U}_E -mesurable si f vérifie l'une des propriétés équivalentes, pour tout a dans $\overline{\mathbb{R}}$.

- i) $\{e ; e \in E, f(e) > a\} \in \mathcal{U}_E$
- ii) $\{e ; e \in E, f(e) \geq a\} \in \mathcal{U}_E$

On note $\mathcal{H}(\mathcal{U}_E)$ la classe des fonctions \mathcal{U}_E -mesurables.

L'étude de $\mathcal{M}(U_E)$ montre qu'une condition nécessaire et suffisante de U_E -mesurabilité pour une fonction f définie sur (E, \mathcal{E}) est qu'elle soit U_E -mesurable et isotone. On a de plus la :

II. 3 - Proposition

- i) $\mathcal{M}(U_E)$ est un treillis σ -complet
- ii) $\mathcal{M}(U_E)$ est un cône convexe.

L'interprétation géométrique de la régression ordinale nécessite d'étudier les variables de carré intégrable. Pour ce faire nous noterons :

$$\mathcal{B} = X^{-1}(U_E) = \{X^{-1}(U) ; U \in U_E\}$$

Le sous σ -treillis de \mathcal{A} induit par U_E . Par suite si $P_{\mathcal{B}}$ est la restriction de P à \mathcal{B} nous notons :

$$L^2(\mathcal{B}) = L^2(\Omega, \mathcal{B}, P_{\mathcal{B}})$$

Le problème posé en régression isotone est de trouver l'élément $m \circ X$ de :

$$\mathcal{H} = \{m \circ X ; m \in \mathcal{M}(U_E)\} \cap L^2(\mathcal{A})$$

le plus proche (en norme $L^2(\mathcal{A})$) de Y . Sa détermination est fondée sur le

II. 4 - Théorème ([2] th. 7.1)

- i) $L^2(\mathcal{B}) = \mathcal{H}$
- ii) $L^2(\mathcal{B})$ est un cône convexe fermé et un treillis σ -complet dans $L^2(\mathcal{A})$.

Ce théorème montre que la construction du meilleur approximant de Y en norme $L^2(\mathcal{A})$ peut être entreprise suivant deux voies. L'utilisation de

la structure de treillis σ -complet de \mathcal{H} conduit à une approche algébrique qui est en général ignorée dans la littérature. Nous montrons au chapitre III que l'approche de GUTTMAN de l'analyse des données ordinales reçoit, en ces termes le fondement théorique qui lui manquait.

La seconde approche utilise une argumentation géométrique fondée sur la structure de cône convexe de \mathcal{H} . Nous exposons ci-dessous ses implications statistiques.

Les propriétés fondamentales d'un projecteur dans l'espace de HILBERT $L^2(\mathcal{A})$ muni du produit scalaire :

$$\langle X, Y \rangle = \int_{\Omega} X(\omega) Y(\omega) P(d\omega) = E(XY)$$

associé à la norme $N_p^2(X) = \langle X, X \rangle$, permet de poser en

II.5 - Définition

Soit Y un élément de $L^2(\mathcal{A})$ et \mathcal{B} un sous σ -treillis de \mathcal{A} ; On appelle Espérance conditionnelle de Y par rapport à \mathcal{B} l'unique élément de $L^2(\mathcal{B})$ qui minimise dans $L^2(\mathcal{B})$ l'intégrale :

$$\int_{\Omega} (Y(\omega) - V(\omega))^2 P(d\omega)$$

Nous notons cet élément $E(Y/\mathcal{B})$.

Remarque

De même que dans le cas usuel du conditionnement par rapport à une sous tribu l'espérance conditionnelle par rapport à un σ -treillis peut être définie dans un cadre plus général que celui de la définition II.5, en termes de dérivée de RADON - NIKODYM ([2], ch. VII-4).

Notre étude est limitée aux propriétés immédiates de l'opérateur $E(./\mathcal{B})$ défini sur $L^2(\mathcal{A})$. Il est a priori non linéaire et c'est un projecteur sur un sous espace convexe fermé de $L^2(\mathcal{A})$.

II. 6 : Proposition

Soient Y, Y_1, Y_2 trois éléments de $L^2(\mathcal{A})$ et k un réel non négatif.
Si \mathcal{B} est un sous σ -treillis de \mathcal{A} :

i) $E(\cdot/\mathcal{B})$ est positivement homogène
$$E(kY/\mathcal{B}) = k E(Y/\mathcal{B})$$

ii) $E(\cdot/\mathcal{B})$ est contractant
$$\|E(Y_1/\mathcal{B}) - E(Y_2/\mathcal{B})\|_2 \leq \|Y_1 - Y_2\|_2$$

iii) $E(\cdot/\mathcal{B})$ est invariant par isométrie :
$$T E(Y/\mathcal{B}) = E(TY/\mathcal{B})$$
 pour toute isométrie T
définie sur $L^2(\mathcal{A})$

iiii) $E(\cdot/\mathcal{B})$ est strictement monotone en tout point Y_1 de $L^2(\mathcal{A})$
$$\langle Y_1 - Y_2, E(Y_1/\mathcal{B}) - E(Y_2/\mathcal{B}) \rangle > 0$$

si $E(Y_1/\mathcal{B}) \neq E(Y_2/\mathcal{B})$

De plus il résulte du fait que la meilleure approximation de Y sur le cône convexe fermé $L^2(\mathcal{B})$ appartient à $E(Y/\mathcal{B})$, la

II. 7 - Proposition.

Soient Y un élément de $L^2(\mathcal{A})$ et \mathcal{B} un sous σ -treillis de \mathcal{A} .
Une fonction Z \mathcal{B} -mesurable appartient à $E(Y/\mathcal{B})$ si et seulement si :

i) $E\{(Y-Z)Z\} = 0$

ii) $E\{(Y-Z)V\} \leq 0$ pour tous V dans $L^2(\mathcal{B})$

iii) $E\{(Y+a E(Y/\mathcal{B})/\mathcal{B})\} = (1+a)E(Y/\mathcal{B})$ pour tout réel a supérieur ou égal à -1 .

Ces deux propositions démontrées dans [2] sont de plus complétées par le corollaire de (II.7 ii)) valable puisqu'ici les constantes sont dans $L^2(\mathcal{B})$:

II. 8 - Corollaire :

Soient Y un élément de $L^2(\mathcal{G})$ et \mathcal{B} un sous σ -treillis de \mathcal{G} .
 Tout élément Z de $E(Y/\mathcal{B})$ vérifie l'égalité :

$$\int_{\Omega} Y(\omega) P(d\omega) = \int_{\Omega} Z(\omega) P_{\mathcal{B}}(d\omega)$$

On dit que l'opérateur $E(./\mathcal{B})$ est invariant en moyenne sur $L^2(\mathcal{G})$.

Les propositions rappelées ci-dessus énoncent les propriétés fondamentales de l'opérateur $E(./\mathcal{B})$. Inversement une caractérisation de l'espérance conditionnelle par rapport à un sous σ -treillis, dans $L^2(\mathcal{G})$ a été donnée par DYKSTRA ([2] pp. 325).

II. 9 - Théorème

Soit (Ω, \mathcal{G}, P) un espace probabilisé et T un opérateur positivement homogène, isotone, invariant en moyenne, idempotent et contractant sur $L^2(\mathcal{G})$, strictement monotone en 0.

Soit $\mathcal{B} = \{TY > a\}; Y \in L^2(\mathcal{G}), a \in \mathbb{R}\}$

On a alors : $T = E(./\mathcal{B})$

§2 - L'INDICE DE REGRESSION ISOTONE.

Soit $\{X_i\}_{i \in [n]}$ une famille finie de variables aléatoires définies sur (Ω, \mathcal{G}, P) et à valeurs dans $(H, \tilde{\mathcal{R}})$. Notons (E, \mathcal{R}) l'espace H^n muni de la relation de préordre produit \mathcal{R} et $X = (X_1, \dots, X_n)$:

$$X : (\Omega, \mathcal{G}, P) \longrightarrow (E, \mathcal{E})$$

le sous σ -treillis de \mathcal{G} engendré par X est alors défini au paragraphe 1 et l'on peut énoncer la conséquence directe de II.7 et II.8

II. 10 - Proposition

Soit Y un élément de $L^2(\Omega)$. Un élément Z de $L^2(\mathcal{B})$ est égal à $E(Y/\mathcal{B})$ si et seulement si :

- i) $E(Z) = E(Y)$
- ii) $\text{Cov}(Z, V) = \text{Cov}(Y, V)$ pour tout V dans $L^2(\mathcal{B})$
- iii) $\text{Var}(Z) = \text{Cov}(Z, Y)$

Notations

On a noté $\text{cov}(Z, V)$ la covariance de Z et V et $\text{var}(Z)$ la variance de Z . On notera dans la suite $\rho(Z, V)$ le coefficient de corrélation linéaire entre les variables Z et V .

Dans la suite nous noterons par abus de langage $E(Y/\mathcal{B})$ tout représentant cette classe d'équivalence. Nous pouvons alors énoncer le

II. 11 Corollaire

- i) $E(Y + a / \mathcal{B}) = E(Y/\mathcal{B}) + a$ pour vecteur a dans \mathbb{R}^p
- ii) $\text{var}(Y) = E\{(Y - E(Y/\mathcal{B}))^2\} + \text{var}(E(Y/\mathcal{B}))$
- iii) $1 - \rho^2(Y, E(Y/\mathcal{B})) = \frac{E\{[Y - E(Y/\mathcal{B})]^2\}}{\text{var}(Y)}$

Démonstration

La relation i) est une conséquence directe de la proposition II.10 et la décomposition de la variance se démontre de façon classique grâce à la relation d'orthogonalité II.7 i). Une conséquence immédiate de II.10 iii) est que :

$$1 - \rho^2(Y, E(Y/\mathcal{B})) = \frac{\text{Var}(E(Y/\mathcal{B})) [\text{Var}(Y) - \text{Var}(E(Y/\mathcal{B}))]}{\text{Var}(Y) \cdot \text{Var}(E(Y/\mathcal{B}))}$$

d'où la relation iii) en utilisant ii).

Par analogie avec la définition du rapport de corrélation donnée

lorsque \mathcal{B} est une sous tribu de \mathcal{A} nous posons ν

II. 12 - Définition

On appelle indice de régression isotone de Y par rapport à $\{X_i\}_{i \in [n]}$ le nombre positif ν tel que :

$$1 - \nu^2 = \frac{E\{[Y - E(Y/\mathcal{B})]^2\}}{\text{var}(Y)}$$

Remarques

On a par définition $\nu = |\rho(Y, E(Y/\mathcal{B}))|$. il en résulte que de même que ν^2 , ν^2 est un coefficient non symétrique et en cas d'ambiguïté nous distinguerons $\nu_{Y/X}$ de $\nu_{X/Y}$. De plus il résulte de II.11.ii) que l'on a toujours :

$$\text{Var}(E(Y/\mathcal{B})) \leq \text{var}(Y)$$

avec égalité si et seulement si $Y = E(Y/\mathcal{B})$ au sens de $L^2(\mathcal{A})$. Alors on a une "dépendance isotone" entre Y et X et $\nu = 1$. Cette notion correspond à la notion de "positive dépendance" introduite par LEHMAN (Testing Statistical Hypothesis -Wiley - 1959).

II. 13 - Proposition

- i) ν varie dans $[0, 1]$
- ii) $\nu = 0$ si et seulement si tout élément non dégénéré de $L^2(\mathcal{B})$ a une corrélation linéaire négative avec Y . Alors $E(Y/\mathcal{B})$ est la variable dégénérée $E(Y)$.
- iii) S'il existe un élément V de $L^2(\mathcal{B})$ non dégénérée et tel que $\rho(Y, V) \geq 0$, alors $E(Y/\mathcal{B})$ réalise

$$\nu = \max \{ \rho(Y, V) ; V \in L^2(\mathcal{B}) \}$$

Démonstration

le i) résulte de la définition de ν . Pour montrer iii) il suffit de remarquer que tout élément Z de $E(Y/\mathcal{B})$ maximise la variance dans

$$\mathcal{C} = \{V; V \in L^2(\mathcal{B}), E(V) = E(Y); \text{cov}(Y, V) = \text{var}(V)\}$$

car par définition Z minimise $E\{(Y-V)^2\}$ dans $L^2(\mathcal{B})$ et donc dans \mathcal{C} où

$$E[(Y-V)^2] = \text{var}(Y) - \text{var}(V)$$

Par suite on associe à tout élément W dans $L^2(\mathcal{B})$ tel que :

$$\text{cov}(Y, W) > 0,$$

$$V = E(Y) + k(W - E(W))$$

avec

$$k = \text{cov}(Y, W) / \text{Var}(W)$$

ce qui assure que V est dans \mathcal{C} et l'on vérifie alors que :

$$\rho(Y, Z) = \frac{\text{cov}(Y, Z)}{\sqrt{\text{var}(Y)\text{Var}(Z)}} = \frac{\sqrt{\text{var}(Z)}}{\sqrt{\text{var}(Y)}} \rightarrow \frac{\sqrt{\text{var}(V)}}{\sqrt{\text{var}(Y)}} = \frac{\text{cov}(Y, V)}{\sqrt{\text{var}(Y)\text{Var}(V)}} = \rho(Y, V) = \rho(Y, W)$$

ce qui démontre le iii). Inversement si toute variable non dégénérée de $L^2(\mathcal{B})$ est corrélée négativement avec Y le II.10 iii) assure la positivité de $\text{cov}(Y, E(Y/\mathcal{B}))$ et l'on a donc :

$$\text{var}(E(Y/\mathcal{B})) = \text{cov}(Y, E(Y/\mathcal{B})) = 0$$

$E(Y/\mathcal{B})$ est donc dégénérée et égale à son espérance mathématique $E(Y)$ d'après II. 10 i), et alors $\nu = 0$ d'après sa définition. ■

Cette proposition adaptée d'un résultat de [2] (Corol. E p.320) donne un premier critère d'optimalité de l'indice de régression isotone, mais ne justifie pas totalement son emploi. En effet la propriété d'homogénéité positive de $E(.|\mathcal{B})$ assure l'invariance de ψ par homothétie sur Y (Prop. II.6 i)) et non sur les éléments de $L^2(\mathcal{B})$. De plus le choix du terme normalisateur $\text{Var}(Y)$ n'est fondé que sur une analogie avec le rapport de corrélation.

C'est pourquoi nous généralisons un résultat démontré dans un cadre linéaire, en cartographie, par GILBERT [49], et posons pour tout dans $L^2(\mathcal{B})$:

$$\psi_{K,p}^2(Y,V) = 1 - \frac{E\{Y - kV\}^2}{[\text{var}(Y)]^p [\text{var}(kV)]^q}$$

$$\psi_p^2(Y,V) = \text{Max} \{ \psi_{K,p}^2(Y,V) ; k \in \mathbb{R}_+ \}$$

où p et q sont deux réels vérifiant $p+q = 1$. Il est clair que l'on a

$$\psi_{Y/X}^2 = \psi_1^2(Y, E(Y|\mathcal{B}))$$

II.13 - Théorème

i) Pour toute valeur de p , $\psi_p^2(Y,V)$ est une fonction monotone non décroissante du coefficient $\rho(Y,V)$

ii) S'il existe dans $[0,1]$ un p pour lequel

$$\psi_p^2(Y,V_1) \leq \psi_p^2(Y,V_2) \quad V_1, V_2 \in L^2(\mathcal{B})$$

cette inégalité est alors vérifiée pour tout p dans $[0,1]$.

Démonstration

Le ii) est une conséquence du i) que l'on démontre en remarquant qu'a priori V ne vérifie pas la proposition II.10 i) et donc :

$$E\{(Y-kV)^2\} = \text{var}(Y) + k^2 \text{var}(V) - 2k \text{cov}(V,Y) + [(E(Y)-k E(V))]^2$$

Si l'on pose :

$$\begin{aligned}\lambda &= k [\text{var}(V)]^{1/2} [\text{var}(Y)]^{-1/2} \\ u &= [\text{var}(V)]^{-q/2} [\text{var}(Y)]^{-p/2} E(Y) \\ v &= [\text{var}(V)]^{-\frac{q+1}{2}} [\text{var}(Y)]^{-\frac{p-1}{2}} E(V)\end{aligned}$$

le facteur $\Psi_{K,p}^2(Y,V)$ s'écrit :

$$\Psi_{K,p}^2(Y,K) = 1 - \lambda^{2p} - \lambda^{-2q} + 2\rho(Y,V) \lambda^{p-q} - \lambda^2 v^2 + 2\lambda uv - u^2$$

Cette fonction atteint un extremum en k lorsque :

$$\frac{\partial}{\partial \lambda} \Psi_{K,p}^2(Y,K) = 0$$

c'est-à-dire lorsqu'est vérifiée l'identité :

$$(*) \quad (p-q) \rho(Y,V) = \lambda^{2q} [\lambda v^2 - uv] + p\lambda^{-q} \lambda^{-1}$$

Lorsque λ varie de 0 à $+\infty$ le terme de droite décrit \bar{R} en croissant. Il existe donc une solution unique λ_0 , qui est un maximum de $\Psi_{K,p}^2(Y,V)$ puisque la dérivée seconde de cette fonction par rapport à λ , est négative.

Le théorème est donc démontré car $\frac{d\Psi_p^2(Y,V)}{d\lambda}$ est positif lorsque λ égale λ_0 .

En effet :

$$\frac{d}{d\rho} \Psi_p^2 = \left[\frac{\partial}{\partial \lambda} \Psi_{K,p}^2 \cdot \frac{d\lambda_0}{d\rho} + \frac{\partial}{\partial \rho} \Psi_{K,p}^2(\lambda) \right]_{\lambda_0}$$

et puisque λ_0 annule $\frac{\partial}{\partial \lambda} \Psi_{K,p}^2$, cette équation s'écrit :

$$\frac{d}{d\rho} \Psi_p^2 = \left[\frac{\partial}{\partial \lambda} \Psi_{K,p}^2(\lambda) \right]_{\lambda_0} = 2 \lambda_0^{p-q} \geq 0 \quad \blacksquare$$

Remarques

Ce résultat généralise celui de KRUSKAL [58,d] montrant qu'il est équivalent de Maximiser ψ_0^2 , $\psi_{1/2}^2$ ou ψ_1^2 .

De plus, ce résultat reste valide lorsque, \mathcal{B} étant une sous tribu de \mathcal{A} , on travaille sur le rapport de corrélation. Dans ce dernier cas, notre démonstration, simplifiée, permet de retrouver le résultat de GILBERT lorsque l'on travaille sur les moments non centrés. C'est pourquoi nous ne précisons pas dans l'énoncé que $\rho(Y,V)$ est un coefficient de corrélation.

§ 3 - ALGORITHMES DE REGRESSION ISOTONE.

Les premiers travaux, en régression isotone sont dus à VAN EEDEN [77] qui étudia l'estimation sous contraintes d'isotonie, en termes de maximum de vraisemblance.

On suppose avoir mesuré (Y,X) sur une population Ω de cardinal fini $|\Omega|$. Sur la tribu des parties $\mathcal{P}(\Omega)$ est définie une mesure uniforme et l'on se propose de construire la fonction $\hat{Y} = m \circ X$ meilleure approximation de Y sous la contrainte : m est isotone sur E image de X . \hat{Y} réalise le minimum de :

$$\frac{1}{|\Omega|} \sum_{\omega \in \Omega} |Y(\omega) - \hat{Y}(\omega)|^2$$

De nombreux algorithmes furent proposés lorsque la famille $\{X_i\}_{i \in [n]}$ est réduite à un élément. Nous parlons alors de régression simple. Lorsque \mathcal{R} est une relation de préordre total, l'algorithme proposé par KRUSKAL ([58,a], [2]ch.2) semble le plus performant. Nous en donnons une description due à YOUNG [15].

Pour simplifier l'exposé nous supposons que Ω est indicé de telle sorte que :

$(\omega_i \leq \omega_j) \Rightarrow (X(\omega_i) \leq X(\omega_j))$ Pour tous ω_i, ω_j dans Ω

L'algorithme opère par regroupements en blocs, des éléments ω_j pour lesquels $Y(\omega_j)$ viole l'ordre d'indiciation. On affecte alors au bloc la moyenne arithmétique des $Y(\omega_j)$ prises sur tous les éléments ω_j du bloc. On construit ainsi $\hat{Y}(\omega_1)$ puis $\hat{Y}(\omega_2)$ etc...

Premier pas :

On affecte à $\hat{Y}(\omega_i)$ ($i = 1, 2$) la valeur de $Y(\omega_i)$.

Si $Y(\omega_1) \leq Y(\omega_2)$ ω_1 forme le premier bloc B_1 de valeur $V_1 = Y(\omega_1)$
et ω_2 forme le second B_2 de valeur $V_2 = Y(\omega_2)$

Si non $\hat{Y}(\omega_1) = \hat{Y}(\omega_2) = \frac{1}{2} [Y(\omega_1) + Y(\omega_2)]$. Il n'y a pas de 1er bloc et $\{\omega_1, \omega_2\}$ forment le bloc B_2 de valeur $V_2 = \hat{Y}(\omega_1)$.

s-ième pas :

On affecte à $\hat{Y}(\omega_s)$ la valeur $Y(\omega_s)$.

Si $Y(\omega_s) \leq V_{s-1}$, ω_s forme le bloc B_s . On calcule la (s+1)-ième disparité.

Si non ω_s forme avec B_{s-1} le bloc B_s de valeur :

$$V_s = (Y(\omega_s) + \text{card}(B_{s-1}) \cdot V_{s-1}) / (\text{card}(B_{s-1}) + 1).$$

Il n'y a alors plus de bloc B_{s-1} . Alors ou bien tous les $\{\omega_i, i \leq s\}$ sont agrégés dans B_s ou bien il existe un bloc B_{s-2} auquel on compare le bloc B_s .

Si $V_s \leq V_{s-2}$, B_{s-2} forme le bloc B_{s-2} .

Sinon B_s et B_{s-2} sont agrégés pour former le bloc B_s qui est comparé au bloc B_{s-3} etc...

Quand tous les blocs nécessaires ont été comparés on effectue le (s+1) ième pas.

L'algorithme s'achève au $|\Omega|$ ième pas et alors toutes les $\hat{Y}(\omega_i)$ sont évaluées et il est immédiat que l'opérateur T qui à Y associe \hat{Y} vérifie le théorème de DYKSTRA (th.II. 9). L'indice de régression isotone s'écrit alors :

$$\nu^2 = 1 - \frac{\sum_{\omega \in \Omega} |Y(\omega) - \hat{Y}(\omega)|^2}{\sum_{\omega \in \Omega} |Y(\omega) - \bar{Y}|^2} \quad \text{avec } \bar{Y} = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} Y(\omega)$$

Remarque :

L'étude de la distribution de probabilité de ν^2 est connue sous l'hypothèse de normalité de la variable Y. PRENTICE remarque en effet dans [70] que ν^2 est la statistique de test de l'hypothèse :

H_0 : les $Y(\omega)$ sont indépendants de même loi $N(m, \sigma^2)$
contre l'hypothèse

H_1 : les $Y(\omega)$ sont de loi $N(m(\omega), \sigma^2)$ avec $m(\omega)$ isotone sur Ω lorsque l'on utilise le test de rapport de vraisemblances maximales.

On montre dans [2] que ν^2 est de même loi que :

$$\sum_{k=1}^m \frac{|S_m^k|}{m!} Z_K$$

où Z_K est une variable aléatoire suivant une loi bêta $B(\frac{K-1}{2}, \frac{m-K}{2})$, $m = |\Omega|$ et le terme $\frac{|S_m^k|}{m!}$ est la probabilité de construire K blocs avec l'algorithme ci-dessus lorsque H_0 est vérifiée.

Cependant les S_m^K sont les nombres de stirling de première espèce, difficilement calculables (par récurrence) [2]. L'utilisation de ce résultat en analyse des proximités s'est avéré inefficace car lorsque m est grand ν^2 suit approximativement une loi $B[\frac{1}{2} \log m, \frac{1}{2} (m - \log m)]$ et donc on accepte que H_1 est vérifiée si

$$\nu^2 \leq 1 - \frac{\text{Log } m + \sqrt{\text{Log } m}}{m}$$

Une telle borne est en pratique beaucoup trop large. Ceci est dû tant à l'inadéquation des méthodes statistiques paramétriques, invoquée

par SIBSON [70], qu'au recours abusif à une hypothèse gaussienne.

L'étude de la régression isotone multiple (correspondant à une famille $\{X_i\}_{i \in [n]}$ avec $n > 2$) est à notre connaissance peu développée. L'algorithme de GEBHARDT [48] décrit ci-dessous, montre la complexité d'une approche directe du problème. Ceci est dû à ce que E est alors muni de la relation d'ordre produit (qui n'est que partielle).

Cette remarque nous fonde à présenter une approche indirecte consistant à construire une relation de préordre total "meilleure approximation" de la relation partielle initiale. L'utilisation de l'algorithme de KRUSKAL permet alors de construire une solution.

3.1 - Approche directe

On suppose définie sur Ω une indication (non unique !) telle que si $X = (X_1 \dots X_n)$:

$$(X(\omega_i) \leq X(\omega_j)) \Rightarrow (\omega_i \leq \omega_j) \quad (*)$$

et l'on procède comme précédemment, par regroupements en blocs B_i auxquels sera affectée la valeur :

$$V_i = \frac{1}{|B_i|} \sum_{\omega_k \in B_i} Y(\omega_k) \quad (**)$$

L'explication de l'algorithme nécessite la définition de deux piles $A(s)$ et $D(s)$ qui au pas s contiennent respectivement les éléments $X(\omega_i)$ pour lesquels une valeur $\hat{Y}(\omega_i)$ a et n'a pas été affectée :

$$A(s) = (X(\omega_1) \dots X(\omega_{K-1})), \quad D(s) = (X(\omega_K) \dots X(\omega_n)).$$

Initialisation

Au départ $A(o)$ est vide et $D(o)$ contient $X(\omega)$ pour tous les ω dans Ω rangés de telle sorte que $(*)$ soit vérifiée. Il en sera alors ainsi tout au long de l'algorithme et à la fin de chaque itération on aura $\hat{Y}(\omega_i) \leq \hat{Y}(\omega_j)$ si $\omega_i < \omega_j$ pour tous $X(\omega_i), X(\omega_j)$ dans $A(s)$.

On pose $A(1) = \{X(\omega_1)\}$ et $\hat{Y}(\omega_1) = \dot{Y}(\omega_1)$

ω_1 forme le bloc B_1 de valeur $V_1 = Y(\omega_1)$

Au pas s

(1) Notons $X(\omega_K)$ le dernier élément de la pile $A(s-1)$ et l'on cherche pour $j = K-1$ à $j = 1$ l'élément de $A(s-1)$ pour lequel on a :

- soit $\hat{Y}(\omega_j) \leq Y(\omega_K)$
- soit $X(\omega_j) \leq X(\omega_K)$

Dans le premier cas on pose $\hat{Y}(\omega_K) = Y(\omega_K)$ et ω_K forme le bloc B_s de valeur $V_K = Y(\omega_K)$. $X(\omega_K)$ devient le dernier élément de $A(s)$. On réalise l'itération $s+1$.

Dans le second cas, on réalise les étapes suivantes :

(2) ω_j appartient au bloc B_v auquel on affecte ω_K . B_v prend une nouvelle valeur V_v calculée grâce à (**). On cherche alors si B_v se décompose en deux blocs notes B_s et B_{s-1} suivant l'algorithme de décomposition donné en [48]. Si tel est le cas $X(\omega_K)$ appartient à B_{s-1} et tout (ω_j, ω_p) dans $B_{s-1} \times B_s$ vérifie :

$$X(\omega_j) > X(\omega_p)$$

Les anciens blocs B_{v+1}, \dots, B_{s-1} sont placés en tête de la pile $D(s)$ de façon à respecter la contrainte (*).

(3) On continue de comparer $Y(\omega_K)$ aux blocs de $A(s-1)$ de même qu'en (1). Si l'élément j cherché n'existe pas on fait $A(s) = A(s-1)$ et l'on étudie ω_{K+1} en allant en (1).

(4) Si j existe, on regarde si les blocs B_{s-1} et B_j peuvent être échangés i.e. sont tels que $X(\omega_i) < X(\omega_p)$ pour tous (ω_i, ω_p) dans $B \times B_{s-1}$.

Si oui, B_{s-1} devient le bloc B_j et pour $\alpha = j+1$ à $s-1$ on fait $\alpha = \alpha + 1$ et l'on itère (3). Sinon on va en (2).

La convergence de cet algorithme vers une solution unique est démontrée par GEBHARDT [48].

II. 14 - Théorème

Supposons que $\{X(\omega), \omega \in \Omega\}$ est décomposé en K blocs B_j tels que :

- i) la valeur V_j de B_j est déterminée par (**)
- ii) $X(\omega), \omega \in \Omega$ est ordonné par la relation d'ordre produit sur E .
- iii) Il n'existe aucun Bloc B_j décomposable. id est tel que $B_j = B_{j1}^1 \cup B_{j2}^2$ avec $V(B_{j1}^1) < V_j$ et $X(\omega_r) < X(\omega_s)$ pour tout (ω_r, ω_s) dans $B_{j1}^1 \times B_{j2}^2$.

Alors $\hat{V}(\omega) = V_j$ si $\omega \in B_j$ est l'unique solution au problème de régression isotone.

La complexité de cet algorithme en limite l'utilisation pratique à l'étude de famille $\{X_i\}_{i \in [n]}$ pour lesquelles n est faible. Dans ce cas même, le temps d'exécution machine croît très vite avec $|\Omega|$.

3. 2 - Approches indirectes

L'approximation de la relation d'ordre partiel \mathcal{R} définie sur E par une relation d'ordre total $\hat{\mathcal{R}}$ est un domaine très étudié en analyse multicritère ([19], [40]). Nous présentons deux méthodes utilisables en régression isotone. La première présentée par de CANI [42, a, b] pour étudier les comparaisons par paires, est transposée dans notre contexte car elle permet de définir $\hat{\mathcal{R}}$ comme l'estimateur de maximum de vraisemblance de \mathcal{R} . Elle est donc conceptuellement homogène aux techniques de régression isotone.

Cependant, l'inefficacité des tests statistiques sur χ^2 , signalée plus haut, pose le problème de la signification des solutions construites par régression isotone multiple. c'est pourquoi nous présentons une méthode algébrique due à JACQUET-LAGREZE [32]. Son but est de rechercher la relation d'ordre $\hat{\mathcal{R}}$ qui maximise un indice de cohésion dont l'interprétation statistique semble intéressante.

METHODE DE MAXIMUM DE VRAISEMBLANCE.

$(X_1, \dots, X_K, \dots, X_n)$ est interprété comme un n-échantillon d'une variable aléatoire X définie sur $(\Omega, \mathcal{G}, \mathcal{P})$ où Ω est de cardinal fini m . On note alors :

$$p_{ij} = P\{X(\omega_i) < X(\omega_j)\}$$

où ζ est la relation de préordre total définie sur chaque coordonnée de E .

$$q_{ij} = P\{X(\omega_i) = X(\omega_j)\}$$

$$p_{ij} = 1 - p_{ij} - q_{ij}$$

Si n_{ij} est le nombre d'observations $(X_K(\omega_i) < X_K(\omega_j))$ et t_{ij} celui des observations $\{X_K(\omega_i) = X_K(\omega_j)\}$, on a :

$$n_{ij} = n - n_{ij} - t_{ij}$$

Et la fraction de vraisemblance d'une observation s'écrit :

$$L(p_{ij}, q_{ij}) = p_{ij}^{n_{ij}} q_{ij}^{t_{ij}} (1 - p_{ij} - q_{ij})^{n - n_{ij} - t_{ij}}$$

Si l'on impose à tout ordre total sur Ω de vérifier $p_{ij} \geq p_{ij}$ pour tout classement tel que ω_i soit classé avant ω_j , de CANI a démontré le :

II. 15 - Théorème

Si ω_i est classé avant ω_j dans une indication de Ω , les estimateurs de maximum de vraisemblance de p_{ij} et q_{ij} sont donnés par :

$$\hat{q}_{ij} = \frac{t_{ij}}{n} \quad \hat{p}_{ij} = \text{Max} \left\{ \frac{n_{ij}}{n}, \frac{n_{ij} + n_{ij}}{2n} \right\}$$

La contribution de la comparaison de ω_i à ω_j au logarithme de la fonction de vraisemblance est donc estimé par :

$$C_{ij} = n_{ij} \log \hat{p}_{ij} + t_{ij} \log \hat{q}_{ij} + (n - n_{ij} - t_{ij}) \log (1 - \hat{p}_{ij} - \hat{q}_{ij})$$

Une indication de Ω est un élément σ du groupe des permutations \mathfrak{S}_m et si $(i,j) \in \sigma$ signifie que ω_i est classé avant ω_j dans la permutation σ , La détermination de l'élément optimal σ dans \mathfrak{S}_n peut être formulée en termes de maximisation de la fonction de vraisemblance. Par monotonie de la fonction logarithmique ce problème équivaut au calcul de :

$$Z_0 = \text{Max} \left\{ \sum_{(i,j) \in \sigma} C_{ij} ; \sigma \in \mathfrak{S}_n \right\}$$

La solution d'un tel problème se construit de façon simple ([42,b]) par un algorithme Branch and Bound car Z_0 est majoré par

$$\bar{Z} = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{Max}(C_{ij}, C_{ij})$$

et donc si $\bar{Z}(ij)$ est un majorant de Z_0 quand ω_i est classé avant ω_j on a :

$$\bar{Z}(i,j) = \bar{Z} - \text{Max}(C_{ij}, C_{ij}) + C_{ij}$$

et de même si $\bar{Z}(i,j,k)$ est un majorant de Z_0 quand ω_i et ω_k sont classés dans cet ordre on a :

$$\bar{Z}(i,j,k) = \bar{Z}(i,j) - \text{Max}(C_{ik}, C_{ik}) - \text{Max}(C_{jk}, C_{jk}) + C_{ik} + C_{jk}$$

$$\bar{Z}(i,k,j) = \bar{Z}(i,j,k) + C_{kj} - C_{jk}$$

$$\bar{Z}(k,i,j) = \bar{Z}(i,k,j) - C_{ik} + C_{ki}$$

et de façon générale si $\omega_{\sigma(1)}, \dots, \omega_{\sigma(\ell)}$ sont classés dans cet ordre ($\ell \leq m$) pour σ , on a

$$\bar{Z}(\sigma(1) \dots \sigma(\ell)) = \bar{Z}(\sigma(1) \dots \sigma(\ell-1)) - \sum_{k=1}^{\ell-1} \text{Max}[C_{k\ell}, C_{\ell k}] + \sum_{k=1}^{\ell-1} C_{k\ell}$$

$$\bar{Z}(\sigma(1) \dots \sigma(\ell), \sigma(\ell-1)) = \bar{Z}(\sigma(1), \dots, \sigma(\ell-1), \sigma(1)) - C_{\ell-1, \ell} + C_{\ell, \ell-1}$$

On peut donc utiliser un algorithme Branch and Bound qui construira la relation d'ordre total cherchée ou de façon équivalente la permutation σ sur Ω . L'application de l'algorithme de KRUSKAL est alors possible et conduit en pratique à une solution beaucoup plus économique que celle de GEBHARDT pourvu que soit choisi une permutation σ_0 convenable pour initialiser l'algorithme Branch and Bound (C.F. [42,b]).

METHODE DE MAXIMUM DE COHESION

Cette démarche, due à JACQUET-LAGREZE [32] construit une relation d'ordre total $\hat{\mathcal{R}}$ dont l'interprétation est particulièrement pertinente en psychométrie, bien qu'il ne soit pas une meilleure approximation de \mathcal{R} au sens des moindres carrés.

II. 16 - Définition

Soit X_0 une variable aléatoire définie sur $(\Omega, \mathcal{G}, \mathcal{P})$ où $|\Omega| = m < \infty$ à valeur dans un espace préordonné (H, \mathcal{R}_H) . On appelle signature de la variable X_0 sur Ω la fonction σ définie $\Omega \times \Omega$ par

$$\begin{aligned} \sigma(\omega_i, \omega_j) = \sigma(i, j) &= 1 \quad \text{si } X_0(\omega_i) \text{ est classé avant } X_0(\omega_j) \text{ dans le préordre } \mathcal{R}_H \\ &= 0 \quad \text{si } X_0(\omega_j) \text{ est classé avant } X_0(\omega_i) \text{ dans le préordre } \mathcal{R}_H \\ &= \frac{1}{2} \quad \text{si } X_0(\omega_i) \text{ et } X_0(\omega_j) \text{ appartiennent à la même classe} \\ &= 0 \quad \text{si } X_0(\omega_i) \text{ et } X_0(\omega_j) \text{ sont non comparables ou non comparés.} \end{aligned}$$

Lorsque l'on étudie une famille $\{X_i\}_{i \in [n]}$ de variables aléatoires définie sur $(\Omega, \mathcal{A}, \mathcal{P})$ l'élément $X : (X_1, \dots, X_n)$ à valeur dans l'espace produit (E, \mathcal{A}) permet de définir la permutation $t (t \in \mathcal{G}_m)$ cherchée lorsque l'on suppose connue une pondération $\{p_K, K = 1, 2, \dots, n\}$ des variables X_K . On note σ_K la signature de X_K .

II. 17 - Définition [32]

Si l'on note A_{ij} l'intensité du classement $(i, j) \in t$:

$$A_{ij} = \sum_{K=1}^n p_K \sigma_K(i, j)$$

On appelle COHESION de la famille $\{X_K\}_{K \in [n]}$, la quantité

$$C_0(t) = \frac{\sum_{(i,j) \in t} \{A_{ij} - A_{ji}\}}{\sum_{i,j} A_{ij}}$$

Remarque

Si l'on note $\zeta_{K,t}$ le coefficient de corrélation de KENDALL entre l'ordre induit par X_K et la permutation t , on remarque qu'en l'absence d'exaequos on a :

$$C_0 = \frac{1}{n} \sum_{K=1}^n \zeta_{K,t}$$

cet indice varie donc dans $[0, 1]$ et mesure la corrélation de rang moyenne de $X_K(\omega_i)$ à la permutation t .

L'algorithme de JACQUET LAGREZE consiste à travers la permutation t (dite optimale) qui dans \mathcal{G}_m maximise $\sum_{(i,j) \in t} A_{ij}$, et donc de façon équivalente $C_0(t)$.

Il est fondé sur le :

II. 18 - Théorème

Si la permutation t induit sur Ω une relation d'ordre et si l'on note $e_t(\omega_1, \omega_p)$ l'écart de ω_1 à ω_p pour t :

$$e_t(\omega_1, \omega_p) = \sum \{ (A_{ij} - A_{ji}) ; (\omega_1, \omega_j) \in t \quad i \geq 1, j \leq p \}$$

Alors la trace de t sur tout sous ensemble d'éléments adjacents (au sens de ℓ_t) est optimale si et seulement si t est optimale.

L'algorithme travaille donc en construisant la trace de t sur une suite de sous tableaux de $A = \{A_{ij} ; 1 \leq i, j \leq m\}$ d'ordre croissant. C'est un algorithme de permutation dont l'exposé nécessite l'introduction d'un langage de théorie des graphes, qui est hors de notre propos. Nous renvoyons donc à l'article de JACQUET LAGREZE paru dans le n°4 de la revue " Informatique en Sciences Humaines"(1969)

§ 4 - LE MODELE LINEAIRE ORDINAL

Notre étude développe dans les chapitres suivants des méthodes de statistique purement descriptives. Cependant le formalisme de régression ordinaire permet de construire un modèle description-prédiction dont nous traçons les grandes lignes.

Sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathcal{P})$ sont définies

i) une variable aléatoire ordinaire X à valeurs dans l'ensemble ordonné (E, \mathcal{E})

ii) un vecteur aléatoire Y à valeurs dans $(\mathbb{R}^m, \mathcal{B}_{\mathbb{R}^m})$ de carré intégrable. On note $Y \in L^2_{\mathbb{R}^m}(\Omega)$ et l'on suppose Y centrée

On suppose qu'existe un élément $\langle \cdot, w \rangle$ dans le dual de $L^2_{\mathbb{R}^m}(\Omega)$ tel que

$$Z(\omega) = \langle Y(\omega), w \rangle = \sum_{K=1}^m w_K Y_K(\omega)$$

factorise de façon isotone à travers X $Z = m \circ X$, m isotone sur (E, \mathcal{E}) .

On se propose de construire la meilleure approximation $\langle \cdot, \hat{w} \rangle$ de $\langle \cdot, w \rangle$ au sens des moindres carrés, sous cette contrainte d'isotonie.

Pour ce faire on utilise le fait que $\hat{Z} = \langle Y, \hat{W} \rangle$ maximise l'indice de régression :

$$v_{Z/X}^2 = 1 - \frac{E\{(Z-V)^2\}}{E(Z^2)} \quad \text{sur les éléments } V \in L_{\mathbb{R}}^2(\mathcal{B})$$

puisque, Y étant centrée, il en est de même de Z.

En pratique, Ω est de cardinal fini p et \mathcal{A} est la tribu de ses parties $\mathcal{P}(\Omega)$. La probabilité P est définie par :

$$P(\{\omega\}) = p_{\omega} \quad \text{avec en général } p_{\omega} = \frac{1}{p}$$

$L_{\mathbb{R}}^2(\mathcal{P}(\Omega))$ est alors identifié à \mathbb{R}^p et l'on minimise :

$$1 - v_{Z/X}^2 = \frac{\sum_{\omega \in \Omega} (Z(\omega) - V(\omega))^2}{\sum_{\omega \in \Omega} Z_{\omega}^2} = \frac{t_{(Z-V)}(Z-V)}{t_{ZZ}} = \frac{t_W t_{YYW} + t_{VV} - 2t_W t_{YV}}{t_W t_{YYW}}$$

sous la contrainte $\text{var}(Z) = 1$, en écrivant le lagrangien

$$L(W, \lambda) = t_W t_{YYW} + t_{VV} - 2t_W t_{YV} - \lambda \{t_W t_{YYW} - 1\}$$

dont la dérivée s'annule si :

$$(1) \quad \begin{cases} (1-\lambda) t_{YYW} = t_{YV} \\ t_W t_{YYW} = 1 \end{cases}$$

Système que l'on résoud en λ et W :

$$\begin{cases} \lambda = 1 - [t_V t_Y P_Y V]^{1/2} \\ W = [t_V t_Y P_Y V]^{1/2} P_Y V \end{cases}$$

où l'on note P_Y le projecteur aux moindres carrés $P_Y = (t_{YY})^{-1} t_Y$

Le problème est résolu si l'on connaît l'élément v qui maximise v^2 dans $L^2(\mathcal{B})$. Il résulte des paragraphes précédents que l'on doit poser $V = \hat{Z} = E(Z/\mathcal{B})$. La solution cherchée est donc :

$$(2) \hat{W} = (\hat{Z}^T t_Y P_Y \hat{Z})^{1/2} P_Y \hat{Z}$$

ALGORITHME DE RESOLUTION

L'expression des équations stationnaires (1) montre que l'on doit utiliser une méthode itérative puisque la contrainte dépend de Y .

o) Initialisation

Si Y est une variable centrée. On peut prendre $\hat{Z}^{(0)} = Y$, le centrage important puisque l'on a supposé $E(X) = 0$. Dans le cas contraire, on prendra pour $Z^{(0)}$ toute fonction monotone, centrée de Y . Un choix optimal est discuté au chapitre III.

Itération s

- (1) Calcul de $w^{(s)}$ par (2) en posant $\hat{Z} = \hat{Z}^{(s-1)}$
- (2) Calcul de $Z^{(s)} = X w^{(s)}$
- (3) Calcul de $\hat{Z}^{(s)}$ par regression isotone et calcul de $v^{2(s)}$
- (4) Si $s = s_{MAX}$ aller à ALGO2
sinon $s = s+1$ Aller en (1).

Dans cet algorithme la valeur de v^2 décroît en (1) et en (3) d'après la proposition II.10. De plus

$$(v^{(s+1)} = v^{(s)}) \Leftrightarrow (w^{(s+1)} = w^{(s)}) \Leftrightarrow (\hat{Z}^{(s+1)} = \hat{Z}^{(s)} = Z^{(s)}) \Leftrightarrow (v^{(s)} = 1)$$

Cependant la convergence est en général lente si $v_{MAX} < 1$.

C'est pourquoi de LEEUW [33,b] propose d'assurer la convergence en un nombre fini d'itérations d'utiliser un 2ème algorithme lorsque $s \gg s_{max}$ - i.e. lorsque l'on est proche de l'optimum. v^2 est alors dominé par son développement de TAYLOR au second ordre et est donc quasi-

quadratique. La méthode de NEWTON RAPHSON est alors très efficace et permet de plus de donner une information sur la distance à l'optimum. Nous reprenons l'exposé de de LEEUW, dans notre formalisme :

$$\begin{aligned} E &= {}^t Y Z = {}^t Y Y W \\ F &= {}^t Y (Z - \hat{Z}) \\ N &= {}^t (Z - \hat{Z}) (Z - \hat{Z}) \quad \Rightarrow \frac{\partial}{\partial w_K} N = 2F_K \\ D &= {}^t Z Z \quad \Rightarrow \frac{\partial}{\partial w_K} D = 2 E_K \end{aligned}$$

On a donc :

$$\frac{\partial}{\partial w_K} (1 - v^2) = \frac{2}{D} (F_K - (1 - v^2) E_K)$$

Pour K-ième composante du gradient $G = \frac{\partial}{\partial w} (1 - v^2)$ et

$$\frac{\partial}{\partial w_K} \frac{\partial}{\partial w_L} (1 - v^2) = \frac{2}{D} (v^2 ({}^t Y Y)_{KL} - [G_L E_K + G_K E_L])$$

Comme terme générique de la matrice Hessienne H. Le schéma de NEWTON-RAPHSON s'écrit alors ([81], [83])

$$(*) \quad W^{(s+1)} = W^{(s)} - [H^{(s)}]^{-1} G^{(s)}$$

et la distance au maximum, qui donne un bon critère d'arrêt des itérations est donnée par :

$$\text{TEST}^{(s)} = v_{\text{MAX}}^2 - [v^2]^{(s)} \approx -\frac{1}{2} {}^t G^{(s)} [H^{(s)}]^{-1} G^{(s)}$$

ALGO 2

- (1) $t = 1, W^{(0)} = w^{(sMAX)}, \hat{Z}^{(0)} = Z^{(sMAX)}, \nu^{(0)} = \nu^{(sMAX)}$.
- (2) calcul de $G^{(t)}, H^{(t)}, [H^{(t)}]^{-1}$ et $TEST^{(t)}$
Si $TEST^{(t)} < \epsilon \Rightarrow STOP$
- (3) calcul de $W^{(t)}$ par (*) puis de $Z^{(t)} = Y W^{(t)}$
- (4) calcul de $\hat{Z}^{(t)}$ par regression isotone puis de $\nu^{(t)}$
- (5) $t = t+1$ Aller en (2)

Remarque :

En pratique, la démarche de de LEEUW présente l'inconvénient de nécessiter l'inversion de la matrice H. Cette difficulté disparaît avec l'utilisation de l'algorithme de variance de DAVIDON [81] qui ne nécessite pas le calcul de H.

Applications du modèle linéaire ordinal

Le modèle exposé ci-dessus permet la régression linéaire simple ou multiple d'une variable quantitative Y par la variable ordinale X uni ou multidimensionnelle.

On peut de même mettre en oeuvre une régression polynomiale ordinale (sans terme constant en considérant $\tilde{Y} = (Y, Y^2, \dots, Y^m)$). Si l'on étudie un polynôme à deux variables indépendantes on prendra $\tilde{Y} = (X_1, X_2, X_1 X_2, X_1^2, X_2^2)$ etc...

Si Y est une variable qualitative à p modalités, on obtient une forme ordinale de l'analyse factorielle discriminante. Suivant le critère d'isotonie, on ne maximise plus la distance intergroupes comme dans le cas linéaire, mais on recherche un recouvrement minimal des groupes puisque l'on impose que les projection YW des groupes soient deux à deux disjointes.

Le modèle d'analyse de la variance isotone, étudié par KRUSKAL [58,b] a été exprimé dans notre formalisme par de LEEUW [33,b]. Le problème est de définir un modèle de mesures conjointes à deux modalités [i.e. un plan bifactoriel, en langage classique.] $\Pi = \Pi_1 \times \Pi_2$ où $\text{card}(\Pi_1) = \ell$ $\text{card}(\Pi_2) = m - \ell$ sur lequel une observation est l'élément $\omega_{ij} = (a_i, b_j)$.

On identifie Π à $\Omega = \{1, 2, \dots, m\}$ et l'on définit sur $(\Omega, \mathcal{P}(\Omega), \mathcal{P})$ le vecteur aléatoire Y à valeurs dans $(\mathbb{R}^m, \mathcal{B}_{\mathbb{R}^m})$ de la façon suivante :

si $\omega_{ij} \in \Omega$, $\omega_{ij} = (a_i, b_j)$ et l'on pose

$$Y_K(\omega_{ij}) = 0 \text{ si } K \neq i \text{ et } K \neq \ell + j$$

$$= 1 \text{ sinon}$$

Par suite on retrouve le modèle additif :

$$Z(\omega_{ij}) = \langle Y(\omega_{ij}), W \rangle = W_i + W_{\ell+j}$$

le modèle permet donc de factoriser Z à travers une variable ordinale mesurée X

$$Z = m \circ X \text{ où } m \text{ est une fonction isotone}$$

généralisant ainsi la problématique de l'analyse de la variance classique. Il est clair que l'on peut traiter de même les plans multi-factoriels. De plus, la structure particulière de Y impose une adaptation de l'algorithme donné à ce problème particulier ([58,b] [33,b], [15]).

§ 5 - ISOTONIE ET PREORDONNANCES

Dans ce paragraphe, nous rappelons quelques définitions et propriétés élémentaires sur les préordonnances et leurs liens avec les indices de distance. Ces résultats, utilisés dans les chapitres suivants, sont empruntés à BENZECRI ([3], [39]) GUTTMAN [52] et DAVY de VIRVILLE [20].

II. 19 - Définition

On appelle préordonnance de degré m sur un ensemble Ω de cardinal fini m une relation de préordre Q définie sur $\Omega \times \Omega$ vérifiant pour tous $\omega_1, \omega_2, \omega_3$ dans Ω .

$$i) Q((\omega_1, \omega_1), (\omega_2, \omega_3))$$

$$ii) Q((\omega_2, \omega_3), (\omega_1, \omega_3)) \Rightarrow (\omega_2 = \omega_3)$$

$$iii) Q((\omega_1, \omega_2), (\omega_2, \omega_1))$$

Remarques :

* La relation $Q(\omega_1, \omega_2), (\omega_2, \omega_1)$ exprime que ω_1 est plus proche de ω_2 que ne l'est ω_3 de ω_4 . Les axiomes traduisent une notion de proximité :

- i) dit que tout point est plus proche de lui-même que de tout autre.
- ii) est une condition de séparation correspondant à l'axiome de positivité d'une distance.
- iii) est une notion de symétrie.

La donnée d'une préordonnance sur I équivaut à la définition d'une relation de préordre sur l'ensemble $\mathcal{P}^2(\Omega)$ des parties à deux éléments de Ω . Ces deux notions sont identifiées dans la suite. On identifiera de même Q à la relation de préordre induite par Q sur l'ensemble B des parties à deux éléments distincts de Ω .

Deux éléments (ω_1, ω_2) et (ω_3, ω_4) de $\Omega \times \Omega$ sont dits exaequos s'ils vérifient $Q((\omega_1, \omega_2), (\omega_3, \omega_4))$ et $Q((\omega_3, \omega_4), (\omega_1, \omega_2))$ et une préordonnance Q n'admettant pas d'exaequos est une ORDONNANCE.

II. 20 - Définition [52]

On appelle SIGNATURE d'une préordonnance Q la fonction σ_Q définie sur $\Omega^2 \times \Omega^2$, à valeurs dans $\{-1, 0, 1\}$ par :

$$\sigma_Q((\omega_1, \omega_2), (\omega_3, \omega_4)) = \begin{cases} +1 & \text{si } Q[(\omega_1, \omega_2), (\omega_3, \omega_4)] \text{ et } \bar{Q}[(\omega_3, \omega_4), (\omega_1, \omega_2)] \\ -1 & \text{si } \bar{Q}[(\omega_1, \omega_2), (\omega_3, \omega_4)] \text{ et } Q[(\omega_3, \omega_4), (\omega_1, \omega_2)] \\ 0 & \text{si } Q[(\omega_1, \omega_2), (\omega_3, \omega_4)] \text{ et } Q[(\omega_3, \omega_4), (\omega_1, \omega_2)] \\ 0 & \text{si } (\omega_1, \omega_2) \text{ et } (\omega_3, \omega_4) \text{ ne sont pas } Q\text{-comparable.} \\ 0 & \text{si } \omega_1 = \omega_2 \text{ et/ou } \omega_3 = \omega_4 \end{cases}$$

Où \bar{Q} est la relation non- Q .

II. 21 Définition

Soit $V = (\mathbb{R}^n, d)$ un espace euclidien et $X : \Omega \rightarrow V$ une représentation de Ω dans V . On appelle préordonnance du système Q_X la structure de préordre de l'ensemble de $m(m-1)/2$ distances entre les points $d_{ij} = d(X(\omega_i), X(\omega_j))$.

Cependant le problème réciproque consistant à comparer la préordonnance d'un système Q_X à une préordonnance de référence est rendu compliqué par l'existence d'exaequos. Chaque auteur a, en la matière, une définition personnelle. Notre définition est strictement équivalente à celle de GUTTMAN [52].

II. 22 - Définition

Une représentation X de Ω dans l'espace euclidien V est une réalisation

- i) FORTE DE Q si la préordonnance du système Q_X vérifie : $\sigma_Q \equiv \sigma_{Q_X}$
- ii) LARGE DE Q si la préordonnance du système Q_X vérifie :

$$(\sigma_Q[(\omega_1, \omega_2), (\omega_3, \omega_4)] = +1) \Rightarrow (\sigma_{Q_X}[(\omega_1, \omega_2), (\omega_3, \omega_4)] = +1)$$

- iii) FAIBLE DE Q si la préordonnance du système Q_X vérifie :

$$(\sigma_Q[(\omega_1, \omega_2), (\omega_3, \omega_4)] = +1) \Rightarrow (\sigma_{Q_X}[(\omega_1, \omega_2), (\omega_3, \omega_4)] \geq 0)$$

Remarque

La définition des exaequos, donnée plus haut, pose quelques problèmes en pratique car elle est adaptée à des données exactes. Lorsqu'elles sont entachées d'erreur, Si l'on est capable d'évaluer une mesure ε de la précision des mesures δ on posera :

$$(\sigma_Q[(\omega_1, \omega_2), (\omega_3, \omega_4)] = 0) \iff (-\varepsilon \leq \delta(\omega_3, \omega_4) - \delta(\omega_1, \omega_2) \leq \varepsilon)$$

Cette attitude, proposée par de LEEUW [33-d], pose le problème du choix de ε .

Il est clair que la donnée d'une indice de distance (Déf. I.8) (resp. d'un indice de dissimilarité (déf. I.6)) définit une ordonnance (resp. préordonnance). Celle-ci ne change pas si l'on substitue à l'indice un nouvel indice, fonction isotone (Def. I.7) du premier. De plus deux indices de distance associés à une même ordonnance sont image l'un de l'autre par une application isotone m telle que $m(0) = 0$.

Nous associons dans la suite un indice de distance privilégié à toute préordonnance. L'indice de rang qui dans le cas d'une préordonnance nécessite une définition à cause du problème des exaequos. Nous utilisons la notion de rang généralisé, due à BESSON [26].

II. 23 - Définition

Soit E un ensemble de cardinal fini n et P une relation de préordre sur E . Et notons pour tout e dans E :

$$\begin{aligned} \text{Maj}(e) &= \{x ; x \in E \quad P(x, e) \text{ et non-}P(e, x)\} \\ \text{Min}(e) &= \{x ; x \in E \quad P(e, x) \text{ et non-}P(x, e)\} \end{aligned}$$

On appelle rang généralisé de e pour P .

$$R(e) = \frac{1}{2} [F(e) - G(e) + n + 1]$$

où $F(e)$ est le rang descendant de e .

$$F(e) = 1 \quad \text{si } \text{Maj}(e) = \emptyset$$

$$= \text{Sup} \{F(x) ; x \in \text{Maj}(e)\} + \text{card} \{F^{-1} [\text{Sup} \{F(x) ; x \in \text{Maj}(e)\}]\}$$

sinon :

et $G(e)$ est le rang ascendant de e

$$G(e) = 1 \quad \text{si } \text{Min}(e) = \emptyset$$

$$= \text{Sup} \{G(x) ; x \in \text{Min}(e)\} + \text{card} \{G^{-1} [\text{Sup} \{G(x) ; x \in \text{Min}(e)\}]\}$$

sinon :

Remarque

$G(e)$ vaut 1 si aucun élément ne lui est strictement inférieur dans P . Sinon $G(e)$ est égal à la plus grande valeur prise par G dans $\text{Min}(e)$, augmentée du nombre d'éléments possédant cette même valeur.

$R(e)$ ainsi construit est d'échelle d'intervalle ce qui n'est pas le cas de la définition classique adoptée pour les rangs. Cette propriété prend une importance conceptuelle très grande en analyse des données ordinales.

Notation

Si Ω est de cardinal n , nous notons p le cardinal de B :
 $p = n(n-1)/2$. On définit alors h l'injection canonique de B dans $\Omega \times \Omega$ et notons $(\omega_1, \omega_2) = h(b)$.

II. 24 - Définition

Soit Q une préordonnance de degré n sur Ω . On appelle indice de rang associé à Q , l'indice de distance R défini sur $\Omega \times \Omega$ par :

$$R(\omega, \omega) = 0 \quad \text{pour tout } \omega \text{ dans } \Omega$$

$$R(b) = \text{rang généralisé de } b \text{ dans } B.$$

II. 25 - Définition [13]

Soit Q une préordonnance de degré n sur Ω . Elle est dite ULTRAMETRIQUE si l'indice de rang associé à Q vérifie :

$$(\text{Max} \{ R(\omega_1, \omega_2), R(\omega_2, \omega_3) \} \leq r) \Rightarrow (R(\omega_1, \omega_3) \leq r)$$

où r est un entier.

Cette définition est justifiée par la :

II. 26 - Proposition

Soit X une réalisation Forte de la préordonnance Q , de degré n , définie sur Ω , dans un espace métrique (M, d) . Une condition nécessaire et suffisante pour que d soit ultramétrique est que la préordonnance Q le soit.

Démonstration

Ceci résulte de la définition de Q :

$$Q[(\omega_1, \omega_2), (\omega_3, \omega_4)] \Leftrightarrow (d_{12} \leq d_{34})$$

et donc dire que d est ultramétrique implique :

$$(d_{13} \leq \text{Max}[d_{12}, d_{23}]) \Leftrightarrow (R(\omega_1, \omega_3) \leq \text{Max}\{R(\omega_1, \omega_2), R(\omega_2, \omega_3)\})$$

CHAPITRE III

ANALYSE DES DONNEES DE PROXIMITE AGREGÉES

Nous étudions ici deux problèmes de représentation de données de proximité dans un espace euclidien. Les six premiers paragraphes sont consacrés aux données GLOBALES primitivement abordées par SHEPARD [68,a].

Nous définissons le problème au premier paragraphe et les théorèmes d'existence de solutions nous conduisent à opposer ce type d'approche à l'analyse hiérarchique.

L'approche heuristique de SHEPARD fait l'objet du paragraphe suivant. Nous nous attachons à en dégager une interprétation statistique en termes d'analyse de la dispersion.

Une reformulation du problème en termes de régression isotone nous permet d'opposer les formulations ultérieures de KRUSKAL [58,a] et de GUTTMAN [52] dans un troisième paragraphe.

Le suivant donne, après une présentation de l'approche de Mc GEE [63,a] en termes d'élasticité, un exposé détaillé de la méthode de l'orthant positif dû à de LEEUW [33,c]

Les algorithmes présentés sont tous itératifs et nécessitent une initialisation. Le paragraphe cinq développe une étude critique des approches actuellement disponibles.

Le sixième paragraphe étudie une généralisation de l'approche de SHEPARD. Ce que nous appelons l'analyse a-dimensionnelle construit une représentation dans un espace de FINSLER sans contrainte de représentation graphique.

L'analyse des données de proximité CONDITIONNELLES est opposée dans le dernier paragraphe, aux méthodes d'analyse de données GLOBALES étudiées précédemment.

Les données de proximité étudiées dans ce chapitre ont en commun d'être AGREGÉES. Le statisticien dispose alors d'un seul tableau de données $\Delta(\omega)$ défini sur $I \times I$ provenant, soit d'une expérimentation unique, soit de l'agrégation de plusieurs mesures expérimentales. Aussi dans ce chapitre noterons nous Δ pour $\Delta(\omega)$.

§1 - LE PROBLEME ETUDIE EN ANALYSE DES DONNEES DE PROXIMITE GLOBALES

II-1 - Définition

Soit δ une application de $I \times I$ dans l'ensemble préordonné (H, \mathcal{Q}) . On appelle préordonnance induite par δ sur I , la relation de préordre \mathcal{Q} définie sur $I \times I$ qui rend δ isotone de $(I \times I, \mathcal{Q})$ sur (H, \mathcal{Q}) .

Le problème étudié en analyse des proximités globales est le suivant :

Sur un ensemble fini de caractères I , est définie une préordonnance φ de degré n . On se propose de construire une application X de I dans un espace euclidien V , r -dimensionnel, de telle sorte que :

- i) La dimension r soit minimale,
- ii) Soient vérifiées de façon "optimale" les contraintes d'isotonie :

$$(1) \quad \sigma_{\mathcal{Q}}(b, b') [d_b - d_{b'}] \geq 0 \quad \forall (b, b') ; (b, b') \in B \times B, \quad \sigma_{\mathcal{Q}}(b, b') \neq 0$$

où d_b est la distance $d(X(i_1), X(i_2))$ si $h(b) = (i_1, i_2)$

Remarques

En pratique la préordonnance Q est induite par une mesure expérimentale δ et la nature de la relation d'isotonie retenue en ii) est liée au traitement réservé aux exaequos.

L'énoncé de la contrainte (1) suppose qu'il n'existe pas de données manquantes. Cette convention sera adoptée dans l'ensemble de notre travail.

L'existence et l'unicité de la solution de ce problème sont discutées ci-dessous. Les résultats énoncés ont été donnés précédemment sans une formulation équivalente et seules les démonstrations nous sont propres.

De nombreux développements théoriques sont nécessaires pour construire des démonstrations rigoureuses dans le cas général ([4], [33-d] et le plus souvent les auteurs "oublent" que δ est d'échelle ordinale pour donner des démonstrations simples. Nous montrons que l'utilisation de l'indice de rang associé à Q (Def. II-25) permet de donner des solutions simples et constructives puisque c'est un indice de distance (Def. I-8).

III-2 - Définition [39]

Soit V un espace euclidien de dimension $n-1$. Un système de points $\{M_i, i = 1, 2, \dots, n\}$ définissent un $(n-1)$ simplexe non dégénéré si les vecteurs $\{\overrightarrow{M_1 M_i}, i = 2, \dots, n\}$ forment une base de V .

Remarque

Lorsqu'il n'y a pas d'ambiguités, nous identifions le point M_i et le vecteur $X(i) = \overrightarrow{OM_i}$.

III-3 - Proposition (SHEPARD [68-a])

Une préordonnance Q définie sur I , de degré n admet une représentation forte dans un espace euclidien V de dimension inférieure ou égale à $n-1$.

Démonstration [39]

Soit V un espace euclidien de dimension $n-1$ et $X = \{X(i), i \in I\}$ un système de points définissant un $(n-1)$ simplexe non dégénéré dans V , il existe alors une forme quadratique q sur V , unique, telle que :

$$d^2(X(i), X(j)) = \overrightarrow{q(X(i) X(j))}, \overrightarrow{X(i) X(j)} \quad (\forall (i,j), (i,j) \in I \times I)$$

car q est totalement déterminée par les $\frac{n(n-1)}{2}$ valeurs $\overrightarrow{q(X(i) X(j))}, \overrightarrow{X(i) X(j)}$ et donc de façon équivalente par les $n(n-1)/2$ valeurs $\overrightarrow{q(X(1) X(i) X(1) X(j))}$ et $\overrightarrow{q(X(1) X(i), X(1) X(i))}$ où i et j varient entre 2 et n ($i \neq j$).

Lorsque le $(n-1)$ simplexe est régulier $d_{ij}^2 = 1$ et q_r est définie positive. L'ensemble des formes quadratiques définies positives sur V formant un cône convexe ouvert, on peut alors construire par continuité une forme quadratique q définie positive à partir de q_r et respectant les relations d'ordre induites par Q . Par suite la matrice Σ associée à q dans la base canonique de V étant décomposable :

$$\Sigma = {}^t T.T,$$

les vecteurs $X^*(i) = T X(i)$ donnent une représentation forte de Q dans V .

III-4 - Lemme [56]

Soit V un espace euclidien de dimension $n-1$. Il existe au plus un point équidistant des sommets d'un $(n-1)$ simplexe non dégénéré.

Démonstration

Supposons qu'il en existe deux. Sans perte de généralité nous pouvons identifier l'un à l'origine et l'autre au vecteur $A = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ a \end{pmatrix}$

Pour tous i, j dans $\{1, 2, \dots, n\}$, nous avons donc en notant $\|\cdot\|$ la norme euclidienne sur V

$$\|X(i)\|^2 = \|X(j)\|^2$$

$$\|X(i)-A\|^2 = \|X(j)-A\|^2$$

La différence membre a membre se simplifie en

$$x_{n-1}^2(i) - (x_{n-1}(i)-a)^2 = x_{n-1}^2(j) - (x_{n-1}(j)-a)^2$$

Soit :

$$a[2 x_{n-1}(i)-a] = a[2 x_{n-1}(j)-a]$$

i et j étant quelconque et le $(n-1)$ simplexe non dégénéré, ceci impose $a=0$.

III-5 - Théorème de HOLMAN [56]

Soit Q une préordonnance de degré n définie sur I . Si Q est ultramétrique, $(n-1)$ dimensions sont nécessaires pour construire une représentation euclidienne forte de Q . On a alors $d_b^2 = r_b$ pour tout b dans B (r_b étant le rang de b dans Q).

Démonstration

La proposition III-3 assure l'existence d'une représentation X de I dans l'espace euclidien $(n-1)$ dimensionnel V .

Notations

Si $V = V_\alpha \oplus V_\beta$ nous noterons $\|\cdot\|_\alpha$ (resp. $\|\cdot\|_\beta$) la norme induite sur V_α (resp. V_β) par $\|\cdot\|$.

Tout vecteur $X = \begin{pmatrix} X_\alpha \\ X_\beta \end{pmatrix}$ de V vérifie alors :

$$\|X\|^2 = \|X_\alpha\|_\alpha^2 + \|X_\beta\|_\beta^2$$

De plus pour tout A dans V , si nous notons $d_{ia} = d(X(i), A)$ pour tout i de I la boule ouverte de centre A et de rayon ρ est

$$B_o(A, \rho) = \{X(i) \mid i \in I ; d_{ia} < \rho\}$$

Nous montrons par récurrence sur n l'existence d'un élément de V tel que :

$$\mathcal{Q}_n) \begin{cases} r_1 : \text{Pour tous } (i, j) \text{ dans } B \text{ } X(i) \text{ et } X(j) \text{ sont équadistants de } A \\ r_2 : \text{si } |I| \geq 2 \text{ pour tous } (i, j) \text{ dans } B \quad d_{ij} > d_{ia} \sqrt{2} \end{cases}$$

Ces propriétés sont immédiates si $n = 1$ ou 2 . Supposons \mathcal{Q}_m vraie pour tout m entier, $m < n$, et notons

$$\rho = \text{Max } [d_{ij} ; (i, j) \in B]$$

Si α est un élément fixé dans I , notons :

$$I_\alpha = \{i ; i \in I \text{ } X(i) \in B_o(X(\alpha), \rho)\} \quad |I_\alpha| = m$$

$$I_\beta = I \setminus I_\alpha$$

Pour tous (i, j) dans $I_\alpha \times I_\alpha$ et tous (k, ℓ) dans $I_\beta \times I_\beta$ est vérifiée

$$(1) \quad d_{ik} \geq \text{Max}(d_{ij}, d_{k\ell})$$

et donc du fait de l'ultramétrie de Q et de l'isotonie forte de la représentation X ,

$$(2) \quad d_{ik} = d_{jk} = d_{i\ell}$$

On peut toujours supposer $m \geq 2$, en intervertissant au besoin les rôles joués par I_α et I_β . d'après (R) I_α admet une représentation forte dans un espace euclidien V_α de dimension $m-1$. On a donc pour tout i dans I_α :

$$X(i) = \begin{pmatrix} X_\alpha(i) \\ 0 \end{pmatrix} \quad X_\alpha(i) \in V_\alpha$$

Sans perte de généralité on peut supposer dans r_1 $A_\alpha = 0$.

Pour tout k dans I_β , $X(k) = \begin{pmatrix} X_\alpha(k) \\ X_\beta(k) \end{pmatrix}$ où $X_\alpha(k) \in V_\alpha$ et la relation (2) s'écrit pour tous (i,j) dans $I_\alpha \times I_\beta$:

$$d_{ik}^2 = d_{jk}^2 \Leftrightarrow \|X_\alpha(i) - X_\alpha(k)\|_\alpha^2 + \|X_\beta(k)\|_\beta^2 = \|X_\alpha(j) - X_\alpha(k)\|_\alpha^2 + \|X_\beta(k)\|_\beta^2$$

tous les éléments de I_α sont donc équidistants de $X_\alpha(k)$ qui, d'après le lemme, est confondu avec $A_\alpha = 0$.

Par suite tout k dans I_β est représenté par $X(k) = \begin{pmatrix} 0 \\ X_\beta(k) \end{pmatrix}$ $0 \in V_\alpha$ et l'hypothèse de récurrence assure que I_β est représentable dans un espace euclidien de dimension $n-m-1$ et donc dans V_β ,

$$X_\beta(k) = \begin{pmatrix} X_\beta(k) \\ \lambda \end{pmatrix} \quad \text{si } m < n-1$$

$$= \lambda \quad \text{si } m = n-1$$

et l'on peut supposer, sans perte de généralité que $A_\beta = \begin{pmatrix} 0 \\ \lambda \end{pmatrix}$ dans V_β . Le théorème sera démontré si l'on vérifie que λ ne peut être nul.

Hors il résulte de (1) et de l'hypothèse r_2 que :

$$(3) \quad d_{ik}^2 \geq d_{kl}^2 > 2 d_{ka_\beta}^2 = 2 \|\tilde{X}_\beta(k)\|_\beta^2$$

$$(4) \quad d_{ik}^2 \geq d_{ij}^2 > 2 d_{ia_\alpha}^2 = 2 \|X(i)\|^2$$

et donc si $\lambda = 0$ on a

$$\|\tilde{X}_\beta(k)\|_\beta^2 = \|X(k)\|^2 \quad \text{et } a_\alpha = a_\beta = 0$$

$X(i)$ et $X(k)$ appartiennent donc à $B_0(0, \frac{d_{iu}}{\sqrt{2}})$ et la forte isotonie de la représentation nous assure de l'ultramétrie de la distance d . Par suite $X(k)$ appartient à $B_0(i, \frac{d_{ik}}{\sqrt{2}})$ et donc à $B_0(i, \rho)$ puisque $\rho \geq d_{ik}$
 k est donc élément de I_α .

On a montré ainsi la vacuité de I_β et par suite ρ n'est pas atteint, ce qui est absurde I étant de cardinal fini. C'est donc que λ ne peut être nul.

Achevons la démonstration en vérifiant (\mathcal{P}) quand $m = n$. Le point $A = \begin{pmatrix} 0 \\ A \end{pmatrix}$ est équidistant des $X(i)$ car l'égalité

$$(5) \quad \|X_\alpha(i)\|_\alpha^2 + a^2 = d_{ia}^2 = d_{ka}^2 = \|\tilde{X}_\beta(k)\|_\beta^2 + (\lambda - a)^2$$

est vérifiée si

$$a = \frac{1}{2\lambda} \{ \|\tilde{X}_\beta(k)\|_\beta^2 - \|X_\alpha(i)\|_\alpha^2 + \lambda^2 \}$$

et alors

$$(6) \quad \begin{aligned} 2 d_{ia}^2 &= d_{ia}^2 + d_{ka}^2 = \|X_\alpha(i)\|_\alpha^2 + \|\tilde{X}_\beta(k)\|_\beta^2 + a^2 + (\lambda - a)^2 \\ &= d_{ik}^2 + 2a(a - \lambda) > d_{ik}^2 \end{aligned}$$

hors il résulte de (6) que

$$\lambda^2 > | \|\tilde{X}_\beta(i)\|_\beta^2 - \|X_\alpha(i)\|_\alpha^2 |$$

et donc

$$0 < a < \lambda$$

Le deuxième point du théorème résulte de ce que l'indice de rang R est un indice de distance et de plus vérifie (1) et (2). l'hypothèse de récurrence implique donc

$$d_{ij}^2 = r_{ij} \text{ pour tous } (i, j) \text{ dans } I_\alpha \times I_\alpha$$

$$d_{k\ell}^2 = r_{k\ell} \text{ pour tous } (k, \ell) \text{ dans } I_\beta \times I_\beta$$

et donc (3) et (4) sont vérifiées par r_{ij} et $r_{k\ell}$ et donc les relations (5) et (6) assurent que $d_{ik}^2 = r_{ik}$ si $(i, k) \in I_\alpha \times I_\beta$ avec

$$\lambda^2 = r_{ik} - \frac{\|X_\alpha(i)\|^2}{\alpha} - \frac{\|\tilde{X}_\beta(k)\|^2}{\beta} .$$

III-6 - Lemme

Soit M une matrice symétrique, doublement centrée, d'ordre n et $\{\lambda_1, \dots, \lambda_n\}$ ses valeurs propres. Soient I et E les matrices Identité et Unité d'ordre n [$E_{ij} = 1 \forall i, \forall j$]. La matrice symétrique, doublement centrée, d'ordre n

$$M_\alpha = M + \alpha \left[I - \frac{E}{n} \right]$$

définie pour tout réel α a mêmes vecteurs propres que M , associés aux valeurs propres

$$\begin{aligned} \mu_i &= \lambda_i + \alpha && \text{si } V_i \neq \mathbf{1} \\ &= 0 && \text{sinon } i = 1, 2, \dots, n \end{aligned}$$

Démonstration

M est doublement centrée, le vecteur $\mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ est donc vecteur propre associé à la valeur propre $\lambda = \mu = 0$ car $M \mathbf{1} = M_\alpha \mathbf{1} = 0$

Si $V_i \neq \mathbf{1}$ est vecteur propre de M associé à la valeur propre λ_i

$$M_\alpha V_i = M V_i + \alpha \left(I - \frac{E}{n} \right) V_i = \lambda_i V_i + \alpha V_i + 0 = (\lambda_i + \alpha) V_i = \mu_i V_i$$

III-7 - Théorème de GUTTMAN

Soit Q une préordonnance de degré n définie sur I . Si Q n'est pas ultramétrique, il existe une représentation forte de I dans un espace euclidien de dimension inférieure ou égale à $n-2$. La distance d_{ij}^2 entre deux points représentatifs des objets i et j dans I est une fonction affine du rang généralisé r_{ij} .

Remarques

Les énoncés de ce théorème donnés précédemment ([59]) sont érronés car ils n'imposent pas la condition de non ultramétrie.

Lorsque Q est réduite à deux classes (i.e. $r_{ii} = 0, r_{ij} = c \neq 0 \forall i, j \in I, i \neq j$) Q est trivialement ultramétrique et le théorème III-6 répond de cette éventualité.

Notre démonstration n'est valide que parce que les rangs généralisés sont d'échelle d'intervalle.

Démonstration

Soit R l'indice de rang associé à Q, M la matrice symétrique doublement centrée

$$M = - \frac{1}{2} (I - \frac{E}{n}) R (I - \frac{E}{n})$$

associée à la matrice symétrique des rangs généralisés. Son terme générique s'écrit :

$$M_{ij} = - \frac{1}{2} [r_{ij} - \frac{1}{n} \sum_{i=1}^n r_{ij} - \frac{1}{n} \sum_{j=1}^n r_{ij} + \frac{1}{n^2} \sum_{i,j} r_{ij}]$$

dite matrice de pseudo produit scalaire. Elle admet les valeurs propres

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k = 0 \geq \lambda_{k+1} \dots \geq \lambda_n$$

Par suite la matrice $M_{\lambda_n} = M - \lambda_n (I - \frac{E}{n})$ est une matrice de Gram de rang $m \leq n-2$. (cf. Lemme IIIⁿ-6) si λ_n est la plus petite valeur propre non nulle de M. Si nous notons $\{E_{ij}\}_{i=1,2,\dots,m}$ les éléments de la base canonique de $j=1,2,\dots,m$ l'espace des matrices d'ordre m à termes réels ([82]) et

$$F_{ij} = E_{ii} + E_{jj} - E_{ij} - E_{ji}$$

Il existe une représentation forte X de I dans V telle que :

(a) $M_{\lambda_n} = X^t X$ puisque c'est une matrice symétrique semi définie positive
 ([82])

(b) $d_{ij}^2 = \sum_{k=1}^m [x_k(i) - x_k(j)]^2 = \sum_{k=1}^m {}^t X_k F_{ij} X_k = \text{Trace} ({}^t X F_{ij} X)$
 $= \text{Trace} (F_{ij} X^t X) = \text{Trace} (F_{ij} M_{\lambda_n})$

Si l'on développe cette expression par linéarité de la trace, en utilisant les propriétés des matrices E_{ij} , il vient :

$$d_{ij}^2 = r_{ij} - 2 \lambda_n (1 - \delta_i^j)$$

où δ_i^j est le symbole de Kronecker.

Remarques

* X est dite représentation CANONIQUE de Q et si l'on connaît une représentation Y issue de M on peut en déduire X en remarquant que :

$$Y = \Lambda^{1/2} {}^t V \text{ si } M \text{ se décompose en } {}^t V \Lambda V$$

et par suite

$$X^t X = M_{\lambda_n} = {}^t V \Lambda V - \lambda_n (I - \frac{E}{n}) = {}^t V [\Lambda - \lambda_n I] V$$

et donc X est obtenu directement par

$$X = {}^t V [\Lambda - \lambda_n I]^{1/2}$$

* Le calcul de λ_n ne pose aucun problème en pratique. On utilisera à cette fin un algorithme de calcul de la plus petite valeur propre tel celui proposé par SCHWARZ [67].

Les résultats exposés ci-dessus donnent peu de renseignements sur la dimension de l'espace de représentation V. Ceci est dû en partie au fait que l'on exige l'isotonie forte. Cependant puisque tout sous-ensemble de I doit être représenté dans V. La dimension de V ne peut être inférieure à s-1 où s est le nombre d'éléments du plus grand sous-ensemble de I sur lequel la restriction de Q est ultramétrique. La mise en oeuvre d'une telle méthode ne pose aucun problème en pratique.

Plus importante nous semble être l'opposition exprimée par les théorèmes de HOLMAN et GUTTMAN, entre analyse des données de proximité et analyse hiérarchique, suivant le degré de respect de l'inégalité ultramétrique par les données. Le choix en faveur de l'une ou l'autre approche pour analyser un tableau particulier est un problème de décision statistique. A notre connaissance aucun test existant ne peut apporter de réponse. Les tests de H-classificabilité, tels celui de LERMAN [13], sont inadéquats pour deux raisons, tout d'abord ils opposent à l'existence d'une structure H-classifiable, l'absence de structure formulée en termes d'uniformité. De plus, les caractères comparés sont décrits par un ensemble de modalités (ou attributs) et ce dernier point est en général inadapté à l'information recueillie en analyse des proximités.

La notion d'unicité d'une solution n'a pas été étudiée et nous montrons dans quel contexte celle-ci doit être recherchée par la

III-8 - Proposition

L'ensemble $V(I)$ des représentations X de I dans V vérifiant les conditions d'isotonie (1), a une structure de cône (en général non convexe) invariant par transformation du groupe de similitudes.

Démonstration

Les conditions d'isotonie (1) s'écrivent de façon équivalente

$$(1') \quad \sigma_Q(b, b') [d_b^2 - d_{b'}^2] \geq 0 \quad (\forall (b, b'), (b, b') \in B \times B \quad \sigma_Q(b, b') \neq 0)$$

$$\Leftrightarrow \sigma_Q(b, b') \text{ Trace } ({}^t X (F_b - F_{b'}) X) \geq 0$$

$$\Leftrightarrow \text{Trace } ({}^t X G_{bb'} X) \geq 0$$

en suivant les notations utilisées dans la démonstration du théorème III-7 et posant

$G_{bb'} = \sigma_Q(b, b') [F_b - F_{b'}]$, (1) est donc équivalente à (1'')

$$(1''') \text{ Trace } (G_{bb'}, X^t X) \geq 0.$$

Par suite

$$V(I) = \{X ; X \in V^I, \text{ Trace}(G_{bb'}, X^t X) \geq 0, (\forall b, b', (b, b') \in B \times B), \sigma_Q(b, b') \neq 0\}$$

est donc un cône.

$V(I)$ est invariant par translation car si D est une matrice diagonale.

La translation ED est telle que :

$$\text{Trace } (ED G_{bb'}, {}^t(ED)) = 0$$

car pour tous (b, b') dans $B \times B$

$$\left. \begin{array}{l} E F_b = 0 \\ E F_{b'} = 0 \end{array} \right\} \Rightarrow E G_{bb'} = 0$$

et donc si X est élément de $V(I)$ il en est de même de $X + ED$.

$V(I)$ est invariant par rotation. En effet si K est une matrice de rotation

$$\text{Trace } (K^t X G_{bb'}, X^t K) = \text{Trace } ({}^t X G_{bb'}, X^t K K) = \text{Trace } ({}^t X G_{bb'}, X) \geq 0$$

$V(I)$ est invariant par homothétie, car si k est un réel on a

$$\text{Trace } (k X G_{bb'}, k^t X) = k^2 \text{ Trace } ({}^t X G_{bb'}, X) \geq 0.$$

Remarque

L'invariance par translation peut être levée en imposant à la représentation X d'être centrée. Si Y est non centrée on définira donc

$$X = (I - \frac{E}{n}) Y$$

pour lequel

Le barycentre de la représentation est confondu avec l'origine des axes. Cette contrainte est implicitement posée dans tous les algorithmes existants.

L'invariance par rotation ne peut être levée puisqu'elle est due au choix d'une distance euclidienne.

L'invariance par homothétie est levée en imposant une condition de normalisation sur X .

III-9 - Proposition

Si p est le cardinal de B ($p = n(n-1)/2$), l'application :

$$D : V^I \rightarrow \mathbb{R}^{p+}$$

qui à X associe $D(X) = \{d_b, b \in B\}$

a pour image, $D(I)$, un cône convexe fermé (en général non polyédral).

Démonstration

$D(X)$ est convexe par suite de la convexité des normes.

De plus l'application qui à $X^t X$ associe $d_b = \text{Trace}(F_b X^t X)$ est linéaire et donc il existe une matrice M telle que

$$D^2 = M \cdot (X^t X)$$

$X^t X$ est une matrice semi définie positive et décrit donc un cône convexe fermé non polyédral il en est donc de même de $\{d_b^2, b \in B\}$. $D(X)$ ne peut donc être polyédral.

§ II - L'APPROCHE HEURISTIQUE DE SHEPARD

SHEPARD [68,a] construit la représentation optimale X^* par déformations successives de $(n-1)$ -simplexes normalisés.

III-10 - Définition

On appelle $(n-1)$ -simplexe normalisé dans V un simplexe à n sommets dans un espace euclidien $(n-1)$ -dimensionnel, dont le barycentre est l'origine des axes et la longueur moyenne des arêtes est 1.

Pour ce faire, un algorithme en deux phases est mis en oeuvre :

PREMIERE PHASE : construction d'une configuration initiale

Partant d'une configuration de départ formée par les n sommets d'un $(n-1)$ -simplexe régulier normalisé, on construit à la s -ième itération un simplexe auxiliaire non normalisé $Y^{(s)} = \{Y^{(s)}(i), i \in I\}$ en poursuivant deux objectifs :

- assurer le respect des relations d'isotonie (1).
- diminuer la dimension r de l'espace de représentation V .

a) Respect de la relation d'isotonie

La différence algébrique $T_b - T(d_b^{(s)})$ est un indicateur de respect de la relation d'isotonie (1) si $T(d_b^{(s)})$ est définie comme la valeur T_b , de l'indice de distance T associé à Q , telle que le rang de b dans la préordonnance Q soit égal au rang de b dans la préordonnance $Q_X(s)$.

Par suite si $d_b^{(s)}$ est trop grande (resp. trop petite) on a alors $T_b - T(d_b^{(s)}) > 0$ (resp. < 0).

A chaque couple b dans B ($h(b) = (i,j)$) est associé le vecteur correctif

$$\alpha_{ij} = \alpha \cdot \frac{T_{ij} - T(d_{ij}^{(s)})}{d_{ij}^{(s)}} (X^{(s)}(i) - X^{(s)}(j))$$

Le paramètre α module l'importance des corrections faites sur l'ensemble de la configuration $X^{(s)}$.

b) Réduction de la dimension r de V

Celle-ci est réalisée en étirant les arêtes du simplexe $X^{(s)}$ de largeur supérieure à la moyenne et en diminuant celles qui lui sont inférieures, grâce au vecteur correctif

$$\beta_{ij} = \beta \cdot \frac{T_{ij} - \bar{T}}{d_{ij}^{(s)}} (X^{(s)}(i) - X^{(s)}(j))$$

où \bar{T} est la moyenne arithmétique des $n(n-1)/2$ valeurs T_{ij} et β un paramètre de contrôle de la correction.

γ) Algorithme

(0) initialisation $X^{(0)} = (I - \frac{E}{n})$

(1) Construction du simplexe auxiliaire $Y^{(s)}$ par déplacement de $X^{(s)}$:

$$\Delta X^{(s)}(i) = \sum \{ \alpha_{ij} + \beta_{ij} : j = 1, 2, \dots, n \quad j \neq i \}$$

$$Y^{(s)}(i) = X^{(s)}(i) + \Delta X^{(s)}(i)$$

(2) Construction de $X^{(s+1)}$ simplexe normalisé par homothétie sur $Y^{(s)}$.

(3) Test d'arrêt Si positif STOP
 Si négatif $s = s+1$ et aller en (1)

Cet algorithme est totalement défini par la formulation d'un test d'arrêt et une valeur numérique pour α et β . Ce dernier point a été résolu empiriquement par SHEPARD. Trois conclusions s'imposent :

- i) choisir $\beta < \alpha/4$ assure un "bon" respect des contraintes d'isotonie.
- ii) choisir $\alpha < 0,2$ assure une "bonne" convergence de l'algorithme en ce sens que 20 itérations assurent une stabilisation de la configuration : $X^{(20)} \approx X^{(21)}$.
- iii) Plus β est petit et meilleure est la précision mais plus lente est la convergence de l'algorithme.

SHEPARD propose comme test d'arrêt de minimiser la fonction d'écart :

$$L = \frac{1}{p} \sum_{b \in B} (T_b - T(d_b))^2$$

ce qui équivaut, compte tenu des contraintes de III-10 à la minimisation de

$$(*) \quad L_c = \frac{\sum_{b \in B} (T_b - T(d_b))^2}{\sum_{b \in B} (T_b - \bar{T})^2}$$

DEUXIEME PHASE : construction d'une solution

La première phase de l'algorithme ne permet, ni de connaître la valeur de r , ni d'obtenir une visualisation des données. Ces deux buts sont atteints par utilisation de l'analyse en composantes principales.

Cette approche présente deux inconvénients : on définit des axes de représentation dont la signification ne peut être établie dans un contexte ordinal (Prop. III-8). De plus, les projections orthogonales ne respectent pas les relations d'ordre -aussi une configuration de faible dimension nécessite la démarche suivante :

- (1) - Projection dans un sous-espace vectoriel du nuage. $\{(X^{(s)}(i), \frac{1}{n}), i \in I\}$ par analyse en composantes principales.
Si le test d'arrêt sur L est positif STOP
- (2) - Application de la première phase de l'algorithme avec $\alpha = 0,2$ et $\beta = 0$. Puis aller en (1).

La démarche de SHEPARD présente l'inconvénient de travailler sur les indices de distance T isotones par rapport à la préordonnance Q . Toute arithmétique sur T méconnaît le caractère ordinal de cet indice.

De plus, si l'on interprète en termes probabilistes cette démarche dans $(B, \mathcal{P}(B), \lambda)$ où λ est une mesure définie par

$$\lambda(\{b\}) = \frac{1}{p} \quad \text{si } p = |B|$$

deux conclusions s'imposent. L'indice d'écart L_c présente une forte similitude avec l'indice de regression isotone. Pourtant l'algorithme de SHEPARD ne s'exprime pas en termes de regression isotone car $T(d_b)$ n'est dans $L^2(B)$ (cf. Ch. II) que si une représentation forte existe et est atteinte par l'algorithme.

Le point central de la démarche de SHEPARD, en termes de réduction de données est le recours à une maximisation de la variance (dans β) pour diminuer la dimension de V . Cette particularité montre que l'on travaille en termes d'analyse de la dispersion sous la contrainte d'isotonie.

BENNETT [37] utilise les résultats de ([54],[60]) pour montrer que si les n points $X(i)$ sont uniformément distribués dans la boule de rayon a dans $V = (\mathbb{R}^r, d)$, alors la variable aléatoire D distance entre deux points dans cette boule a pour densité de probabilité :

$$f_n(\lambda) = n 2^n \lambda^{n-1} F_{\left(\frac{n+1}{2}, \frac{1}{2}\right)}(1-\lambda^2) \quad \left(\lambda = \frac{d}{2a}\right)$$

où $F_{p,q}(t)$ est la valeur en t de la fonction Béta :

$$F_{p,q}(t) = \frac{1}{\beta(p,q)} \int_0^t u^{p-1} (1-u)^{q-1} du$$

il en résulte ([60]) que D suit asymptotiquement une La Gaussienne de variance

$$\text{Var}(D) = \frac{a^2}{2r}$$

et est donc une fonction décroissante de r (les simulations de BENNETT montre que l'approximation gaussienne est fondée dans les applications pratiques dès que $n \geq 5$). Il en résulte qu'une augmentation de la variance réalisée en assure une diminution de r tant que ceci est compatible avec les conditions d'isotonie.

§ III - FORMULATION DU PROBLEME EN TERMES DE REGRESSION ISOTONE

L'algorithme de SHEPARD se heurte à deux problèmes. Les indices de distance T_b sont d'échelle ordinale et il est donc injustifié d'utiliser leurs valeurs numériques. De plus dans sa formulation initiale SHEPARD utilise directement les indices de dissimilarité δ_b .

De plus la dimension r est inconnue ce qui impose l'utilisation de l'analyse en composantes principales dans la deuxième phase.

Ces deux difficultés sont levées dans les formalisations ultérieures ([58,a] , [52]) étudiées dans ce paragraphe. Le problème est simplifié en supposant r fixé.

On considère deux éléments aléatoires définis sur l'espace probabilisé $(B, \mathcal{G}(B), \lambda)$ (les notations étant celles du paragraphe précédent) :

- i) La donnée Δ , à valeurs dans l'ensemble ordonné (H, \mathcal{R})
- ii) La distance D à valeurs dans l'espace probabilisable $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+})$.

Nous notons \mathcal{B} le sous σ -treillis engendré par Δ ou de façon équivalente par la préordonnance Q définie sur I . X étant une représentation de I dans $V = \mathbb{R}^r$, on recherche l'élément D de $D(I)$ (cf. prop. III-9) et l'élément T dans $L^2(\mathcal{B})$ qui minimisent :

$$\psi = 1 - \nu^2 = \frac{E\{|D-T|^2\}}{\text{Var}(D)}$$

Remarques

Lorsque D est connu, le problème est résolu en termes de régression isotone en prenant $T = E(D/\mathcal{B})$ (cf. Ch. II).

Nous faisons donc une analyse de la dispersion et cherchons (comme en analyse en composantes principales) l'élément D qui maximise (prop. II-11, iii))

$$\nu^2 = \frac{\text{Var}\{E(D/\mathcal{B})\}}{\text{Var } D}$$

Ces méthodes sont d'ordinaire présentées en termes d'adéquation (<< goodness of fit >>) et $1 - \nu^2$ porte alors le nom de STRESSFORM2 (KRUSKAL).

III-3-1 - Interprétation géométrique

B étant ici de cardinal fini p, on identifie de façon classique $L^2(\mathcal{P}(\mathcal{B}))$ à \mathbb{R}^p que l'on munie de la norme euclidienne

$$\|U\|_p = \frac{1}{p} \left[\sum_{b \in B} u_b^2 \right]^{1/2} .$$

$L^2(\mathcal{B})$ s'identifie alors avec le cône \mathcal{C}_0 de \mathbb{R}^p , des indices de distance défini par :

$$\mathcal{C}_0 = \{T ; T \in \mathbb{R}_+^p ; \Sigma T \geq 0\}$$

où Σ est la matrice carrée d'ordre p associée à σ_Q :

$$\Sigma = \begin{pmatrix} 1 & & & & 0 \\ -1 & & & & \\ & 1 & & & \\ & -1 & 1 & & \\ & & \vdots & \ddots & \\ & & & \ddots & 1 \\ & 0 & & & -1 \end{pmatrix}$$

de même, d'après la proposition III-9 D est élément du cône convexe $D(I)$ et le problème consiste à calculer le couple (D^*, T^*) de $D(I) \times \mathcal{C}_0$ réalisant la distance

$$\begin{aligned} d(D(I), \mathcal{C}_0) &= \text{Min} \left[\frac{\|D-T\|_p}{\|D-\bar{D}\|_p} ; D \in D(I), T \in \mathcal{C}_0 \right] \\ &= \text{Min} \{ \text{Min} \left[\frac{\|D-T\|_p}{\|D-\bar{D}\|_p}, T \in \mathcal{C}_0 \right] ; D \in D(I) \} \end{aligned}$$

où \bar{D} est la moyenne arithmétique $\bar{D} = \frac{1}{p} \sum_{b \in B} d .$

III-11 - Proposition

Si $\mathcal{C}_0 \cap D(I) = \{0\}$, alors la distance $d(D(I), \mathcal{C}_0)$ est réalisée par (D^*, T^*) où D^* (resp. T^*) appartient à la frontière de $D(I)$ (resp. \mathcal{C}_0).

Démonstration

\mathcal{C}_0 est d'intérieur non vide s'il existe un couple (b, b') dans $B \times B$ ne réalisant pas l'inégalité ultramétrique.

Supposons donc qu'il existe T appartenant à l'intérieur de \mathcal{C}_0 et réalisant $d(D(I), \mathcal{C}_0)$. Par convexité de \mathcal{C}_0 il existe donc T' appartenant à la frontière de \mathcal{C}_0 et tel que

$$T' = \lambda T + (1-\lambda)D \text{ pour un } \lambda \in]0, 1[$$

et donc

$$\|D-T'\| = \lambda \|D-T\| < \|D-T\|$$

Par suite T ne peut réaliser $d(D(I), \mathcal{C}_0)$.

Le théorème II-13 permet de conclure, en remarquant que le cas d'un élément D' appartenant à l'intérieur de $D(I)$ se traite de façon identique en cherchant

$$\text{le minimum de } \frac{\|D-T\|_p}{\|T-\bar{T}\|_p} \text{ où } \bar{T} = \frac{1}{p} \sum_{b \in B} d_b.$$

La non vacuité de l'intérieur de $D(I)$ est assurée s'il existe une représentation X de I pour laquelle il existe un couple (b, b') dans $B \times B$ tel que $d_b \neq d_{b'}$.

III-12 - Corollaire

Lorsque $\mathcal{C}_0 \cap D(I) = \{0\}$ tout élément X de V^I associé à D^* est une réalisation faible de la préordonnance Q .

En pratique, s'il est aisé de vérifier les conditions de non vacuité de \mathcal{C}_0 et $D(I)$ en recherchant dans les deux cas les couples (b, b') violant l'inégalité ultramétrique, aucun résultat ne nous permet de déterminer quand a tout $\mathcal{C}_0 \cap D(I) = \{0\}$. Il est clair que ceci est lié à la valeur de r (cf. Th. III-6 et III.7) mais aucune approche de ce problème ne nous est connue. L'égalité à 1 de v^2 est, mais nous le verrons, un critère en général fallacieux

d'intersection non réduite à $\{0\}$ de $\mathcal{C}_0 \cap D(I)$.

III- 3-2 - Résolutions algorithmiques : minimisation dans \mathcal{C}_0

Celles-ci sont différenciées par les choix faits pour résoudre le problème :

$$(a) \quad \text{Min} \left\{ \frac{\|D-T\|_p}{\|D-\bar{D}\|_p}, T \in \mathcal{C}_0 \right\}$$

L'approche de KRUSKAL [58,a] a été exposée en II.3 et construit la solution $\hat{T} = E(D/\mathcal{O})$ par un algorithme de regression isotone. Cette construction assure le théorème suivant, du à KRUSKAL [58.c].

III-13 - Théorème

$$\frac{\|D-E(D/\mathcal{O})\|_p}{\|D-\bar{D}\|_p} \text{ est une fonction continuellement différentiable en } D.$$

GUTTMANN [52] propose une méthode de permutation pour construire T^* réalisant (a), sous le nom "d'image de rang". Nous l'exposons tout d'abord lorsque Q est une ordonnance. Notons M (resp. N) la matrice carré d'ordre p de terme générique

$$\begin{aligned} M_b^k &= 1 \text{ si } R(b) = k & \text{resp. } N_b^k &= 1 \text{ si } R_X(b) = k \\ &= 0 \text{ sinon} & &= 0 \text{ sinon} \end{aligned}$$

où R (resp. R_X) est l'application qui a b dans B associe son rang dans l'ordonnance Q (resp. Q_X). Ceci permet de définir la matrice de permutation $\tau = MN$ de terme générique

$$\begin{aligned} \tau_{bb'} &= 1 \text{ si } R(b) = R_X(b') \\ &= 0 \text{ sinon} \end{aligned}$$

et l'on note T^* l'élément de l'intérieur de \mathcal{C}_0 construit par

$$T^* = \tau \circ D(X)$$

T^* offre une alternative aux méthodes de regression isotone, telles que celle de Kruskal. En effet, on vérifie de façon immédiate que la fonction positivement homogène τ n'est pas contractante et viole donc le théorème II.9. Cependant il en résulte que T^* ne vérifie pas la proposition III.11.

Lorsque Q est une ordonnance τ est déterminée de façon unique. Lorsque Q est une préordonnance, on la transforme en préordonnance en posant pour tous (b, b') dans $B \times B$ tels que $Q(b, b')$ et $Q(b', b)$:

- i) non- $Q(b', b)$ si $Q_X(b, b')$ et non- $Q_X(b', b)$
- ii) non $Q(b, b')$ si $Q_X(b', b)$ et non- $Q_X(b, b')$
- iii) un ordre arbitraire si $Q_X(b, b')$ et $Q_X(b', b)$

De toute façon l'imprécision des calculateurs nous assure en pratique de n'être jamais dans l'occurrence iii).

Cette démarche présente l'inconvénient de ne pas construire un T^* dépendant continuellement de $D(X)$ et donc $\frac{\|D-T^*\|_p}{\|D-\bar{D}\|_p}$ n'est pas différentiable en tous points.

L'approche de GUTTMAN est une tentative de rationalisation de celle de SHEPARD qui définit l'indice de distance

$$T_S = \tau \circ D(X)$$

T_S n'appartient à \mathcal{C}_0 que si X est la solution optimale recherchée. Elle trouve de plus sa justification en l'utilisation de la structure de treillis de $L^2(\mathbb{R}) = \mathcal{C}_0$.

III-3-3 - Résolutions algorithmiques : minimisation dans $D(I)$

Les deux applications :

$$V : D(I) \rightarrow \mathcal{C}_0$$

$$D \rightsquigarrow V(D) = T$$

$$U : \mathcal{C}_0 \longrightarrow D(I)$$

$$T \rightsquigarrow U(T) = D$$

existent toujours si aucune contrainte n'est imposée et permettent d'exprimer

$$\psi(X) = 1 - v^2 \frac{D(X)}{T(X)}$$

en fonction de D seul ou de T seul. L'approche de SHEPARD est une tentative de minimisation de

$$\psi(X) = \psi(U(T), T)$$

Alors la relation entre D et X est implicitement exprimée dans U et l'on recherche T via l'ensemble Y des paramètres de la fonction isotone

$$T = m(\Delta)$$

en résolvant l'équation stationnaire

$$\frac{d\psi}{dY} = \left[\frac{\partial \psi}{\partial T} + \frac{\partial \psi}{\partial U(T)} \cdot \frac{dU(t)}{dt} \right] \cdot \frac{dT}{dY} = 0$$

qui permet une minimisation de ψ sous contraintes par l'intermédiaire de U. La formulation de l'équation fondamentale de l'algorithme de SHEPARD.

$$X^{(s+1)}(i) = X^{(s)}(i) + \sum_{j=1}^n [\alpha (T_{ij} - V(d_{ij}^{(s)})) + \beta (T_{ij} - \bar{T})] \frac{X^{(s)}(i) - X^{(s)}(j)}{d_{ij}^{(s)}}$$

apparaît comme une approximation de l'algorithme de gradient :

$$X^{(s+1)}(i) = X^{(s)}(i) - \gamma^{(s)} \left\{ \frac{d\psi}{dY} \right\}^{(s)}$$

dont la faiblesse est d'imposer à α et β d'être fixes pendant tout l'algorithme, devant l'impossibilité de donner l'expression explicite de la fonction m.

Les approches ultérieures étudient la minimisation sans contrainte (exprimées à travers V) de

$$\Psi(X) = \Psi(D, V(D))$$

où la relation entre T et Δ est implicitement exprimée dans V. On est alors amené à résoudre l'équation stationnaire

$$(*) \quad \frac{d\Psi}{dX} = \left[\frac{\partial\Psi}{\partial D} + \frac{\partial\Psi}{\partial V(D)} \cdot \frac{dV(D)}{dD} \right] \cdot \frac{dD}{dX} = 0$$

dont la résolution conduit à un algorithme en une phase dont chaque itération se décompose en deux étapes

- i) Amélioration de la configuration : $D = D^{(s)}$
- ii) Construction de l'indice de distance $T^{(s)} = V(D^{(s)})$.

L'étape ii) est discutée en III.3.2 et l'amélioration de la configuration est réalisée en supposant $T^{(s)}$ fixe, par déplacement de la configuration dans la direction de plus grande décroissance de $\Psi(X)$ donnée par le gradient

$\left[\frac{d\Psi(X)}{dX} \right]^{(s)}$, modulée par un paramètre $\alpha^{(s)}$:

$$X^{(s+1)}(i) = X^{(s)}(i) - \alpha^{(s)} \left[\frac{d\Psi(X)}{dX} \right]^{(s)}(i)$$

où le gradient a pour expression en (i) :

$$\left[\frac{d\Psi(X)}{dX} \right]^{(s)}(i) = \frac{\|D-T\|_p}{\|D-\bar{D}\|} \cdot \sum_{(k,l) \in B} (\delta_i^k - \delta_i^l) \left[\frac{d_{kl} - T_{kl}}{\|D-T\|_p^2} - \frac{d_{kl} - \bar{D}}{\|D-\bar{D}\|_p^2} \right] \frac{X(k) - X(l)}{d_{kl}}$$

où δ_i^k est le symbole de Kronecker.

Cette démarche adoptée par GUTTMAN et KRUSKAL conduit à deux difficultés. Les méthodes de gradient sont sensibles aux minima locaux et d'autre part, si elles approchent très vite de la solution, à son voisinage elles sont incapables le plus souvent de l'atteindre. La stratégie exposée en II-4 utilisant alors une itération de NEWTON RAPHSON est alors souhaitable :

$$X^{(s)}(i) = X^{(s)}(i) - [H^{(s)}]^{-1} \left[\frac{d\Psi(X)}{dX} \right]^{(s)}(i)$$

où $H^{(s)}$ est la matrice Hessienne associée à $\Psi(X)$.

Il est à remarquer que la sensibilité des algorithmes proposés aux minima locaux, n'est pas uniquement dû au choix de méthodes de gradient. Contrairement à ce qui est généralement admis.

En effet le recourt à un algorithme en deux étapes conduit à considérer dans l'étape i) $\Psi(X)$ comme un indice de regression curviligne et non isotone lorsque T est fixé. Ainsi si l'on avait

$$T^{(s)} = V(D^{(s)})$$

il est admissible que l'algorithme construite à l'étape i)

$$D^{(s+1)} = V^{-1}(T^{(s)})$$

conduisant ainsi à un minimum local. Nous nous proposons d'éviter cette difficulté en expérimentant des algorithmes duaux de calcul de la distance entre deux convexes (\mathcal{C}_0 et $D(I)$) tels celui de LAURENT [22].

Une autre solution est proposée par LINGOES et ROSKAM [34] utilisant un algorithme en deux phases

i) Construction de la distance $D^{(s)} = U(T^{(s-1)})$

ii) Construction de l'indice de distance $T^{(s)} = V(D^{(s-1)})$

Par minimisation SANS contraintes de $\Psi(X)$, réalisable en résolvant le système différentiel

$$(\star\star) \quad \left\{ \begin{array}{l} \frac{\partial \Psi(X)}{\partial V(D)} = 0 \\ \frac{\partial \Psi(X)}{\partial U(T)} = 0 \end{array} \right.$$

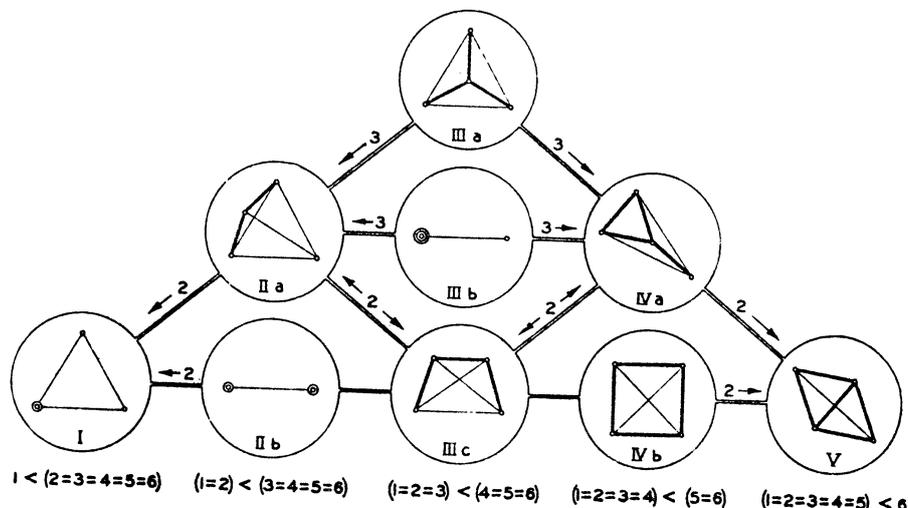
Par des méthodes de gradient (i)) et regression isotone ou permutation (ii)). Une telle stratégie est peu sensible aux minima locaux car $D^{(s)}$ et $T^{(s)}$ sont à chaque itération calculés de façon distincte. Cependant rien n'assure la convergence d'un tel algorithme aussi ROSKAM et LINGOES utilisent dans NIMISSA une stratégie en deux algorithmes

- (1) Minimisation SANS contraintes de $\psi(X)$ grâce à (**)
- (2) Minimisation SOUS contraintes de $\psi(X)$ grâce à (*)

Cette démarche optimale en l'état actuel des recherches algorithmiques, est cependant sensible à deux problèmes que l'on ne peut éviter.

i) La non unicité de la solution (particulièrement fréquente lorsque $\psi(X^*) = 0$) où l'on a introduit une structure dans X qui n'était pas contenue dans les données Δ . On s'en prémunit alors en réalisant plusieurs tentatives algorithmiques.

ii) Les solutions dégénérées sont plus fréquentes et leur occurrence semble mal comprise actuellement. La construction de X ne peut traduire une information qui est dans Δ . Le corollaire III-12 éclaire le problème. Les contre exemples donnés par SHEPARD ([68,a,c]) et BENZECRI [39] sont expliqués par le théorème de HOLMAN (th. III-5). Les dégénérescences sont détectées par l'existence de points confondus $X(i) = X(j)$ où à égale distance $d_b = d_b$. Ils sont donc dus au fait que l'on ne peut construire que des représentations faibles (cor. III.12). Mais contrairement à ([39], [68]) nous pensons que ce qui est fondamental, ce n'est pas l'existence des exactos mais le respect de l'inégalité ultramétrique et l'étude de SHEPARD dans [68,a] trouve une interprétation immédiate en termes de géométrie ultramétrique. Nous nous permettons de reproduire son schéma.



CONFIGURATIONS DEGENEREEES DE 4 POINTS DANS \mathbb{R}^2 . LES 6 DISTANCES NE PRENNENT QUE DEUX VALEURS DISTINCTES

Ainsi à l'indice proposé par SHEPARD pour déterminer le degré de dégénérescence d'une configuration (consistant au rapport du nombre d'exaequos de D au nombre d'éléments de D) préférons nous le rapport du nombre de triplets (i,j,k) pour lesquels l'inégalité ultramétrique est vérifiée dans D à $\frac{n(n-1)(n-2)}{6}$. La comparaison de ces triplets à ceux vérifiant l'inégalité ultramétrique dans Δ est un outil très utile en pratique.

Il importe, en cas de dégénérescence, de traiter les données par des méthodes paramétriques afin de respecter mieux la structure contenue dans les données par l'adjonction d'hypothèses de travail plus fortes que l'isotonie. L'utilisation de méthodes de regression polynomiale proposée par SHEPARD [68,c] se heurte au non respect des relations d'isotonie. Cependant le recourt au modèle linéaire ordinal permet (en théorie) d'assurer l'isotonie dans une telle démarche.

Le problème de quasi dégénérescences doit aussi être mis en lumière. Nous notons ainsi, suivant SHEPARD [68,c] l'existence d'exaequos dans l'indice de distance T^* . Il est clair que l'algorithme de KRUSKAL construit toujours des indices \hat{T} pour lesquels existent des exaequos correspondant à chacun des blocs. Cependant, si \hat{T} est réduit à un petit nombre de blocs, ceci indique que la distance D respecte peu la structure ordinale induite par Δ . Seule l'observation

du diagramme de SHEPARD (le nuage de points (d_b, \hat{T}_b)) est un indicateur de quasi-dégénérescence. Dans un tel cas on utilisera des méthodes paramétriques pour représenter les données, appliquées à l'indice de rang R. L'algorithme de GUTTMANN est une alternative intéressante car l'algorithme de permutation ne construit des T^* quasi-dégénérées que si D est dégénérée.

Remarque

L'interprétation usuelle des cas de dégénérescence ou de quasi dégénérescence en termes d'incapacité des algorithmes à rendre compte de l'information contenue dans les données est le plus souvent abusive. S'il est exact que des données "quasi-ultramétriques" sont redevables de l'analyse hiérarchique, il n'est pas moins vrai que les relations ordinales contenues dans Δ peuvent ne rendre compte d'aucune structure car le résumé exhaustif de Δ est le couple $(\tilde{\Delta}, R)$ et non R : nous avons noté $\tilde{\Delta}$ la statistique d'ordre associée à Δ :

$$\tilde{\delta}_{b_1} \leq \tilde{\delta}_{b_2} \leq \tilde{\delta}_{b_3} \dots \leq \tilde{\delta}_{b_p}$$

§ 4 - LES METHODES DE Mc GEE et de DE LEEUW

Nous exposons ici deux approches qui malgré leur intérêt théorique, n'ont pas retenu l'attention méritée. L'approche de Mc GEE ([35] [63,a]) est fondée sur une argumentation mécaniste et par là est analogue à l'analyse des données usuelles. La démarche de DE LEEUW [33,c] se caractérise par un choix original de l'espace dans lequel le problème est formulé. Elle conduit à de très intéressantes interprétations statistiques.

III-4-1 - Analyse de Mc GEE

On suppose que les éléments i et j dans I sont reliés par un ressort :

- de longueur initiale $\delta_{ij} = \delta_b$
- de longueur finale $d_{ij} = d_b$
- de module d'élasticité $\lambda_{ij} = \lambda_b$

et l'on suppose que pour tout b dans B on a

$$\Delta_b \lambda_b = c^2 \text{ une constante}$$

La configuration initiale est alors contrainte à se projeter dans $V = \mathbb{R}^r$ par extensions et compressions de telle sorte que soit minimisé le travail

$$W = \sum_{b \in B} W_b$$

$$W_b = c^2 \left[1 - \frac{d_b}{\delta_b} \right]^2$$

Cette formalisation de l'approche de SHEPARD conduit à minimiser

$$W = c^2 \left\| 1 - \frac{D}{\Delta} \right\|_p^2$$

Problème résolu par un algorithme de Gradient

$$W^{(s+1)}(i) = X^{(s)}(i) - \alpha^{(s)} \left[\frac{dW}{dX} \right]^{(s)}(i)$$

où

$$\left[\frac{dW}{dX} \right](i) = c^2 \sum_{k,l} (\delta_i^k - \delta_i^l) \left(1 - \frac{d_{ij}}{\delta_{ij}} \right) \frac{\delta_{ij}^2}{d_{ij}^2} \cdot \frac{X(k) - X(j)}{d_{ij}^2}$$

Remarques

* La mise en oeuvre d'une telle démarche se heurte à deux difficultés : la détermination de c^2 et l'utilisation des propriétés numériques de Δ qui est d'échelle ordinale. Ce deuxième point pouvant être corrigé de deux façons :

i) en identifiant Δ au tableau des indices de rang R

ii) en minimisant $\tilde{W} = c^2 \left\| 1 - \frac{R}{D} \right\|^2$ et alors

$$\left[\frac{d\tilde{W}}{dX} \right](i) = c^2 \sum_{k,l} (\delta_i^k - \delta_i^l) \left(1 - \frac{R_{ij}}{d_{ij}} \right) \frac{X(k) - X(j)}{d_{ij}^2} \cdot R_{ij}^2$$

* Pour un même travail W la démarche de Mc GEE conduit à imposer de plus grandes déformations aux grandes dissimilarités qu'aux petites. Les relations d'isotonie locale sont donc privilégiées, ce qui est souhaitable en psychométrie.

III-4-2 - Analyse de DE LEEUW

La méthode de l'orthant positif [33,c] étudie la représentation euclidienne de $A \times A$ données. Nous la présentons ici dans le cas particulier de $B \times B$ données de proximité. Il s'agit d'une méthode de projection très générale qui construit une représentation X de I dans $V = \mathbb{R}^r$ muni d'une distance de Holder L_p , notée d . Nous nous limitons au cas euclidien.

Notations

M est le sous-ensemble de $B \times B$ défini par :

$$((b,b') \in M) \Leftrightarrow ((b,b') \in B \times B \quad \sigma_Q \quad b,b' \geq 0)$$

et si $m = |M|$, on identifie M et $\{1,2,\dots,m\}$.

On notera $k(t)$ l'injection canonique de M dans I^4 :

$$k(t) = (i,j,k,\ell)$$

On associe à tout élément X dans $V(I)$ le vecteur Z dans \mathbb{R}^m défini par

$$Z = \{Z_t ; t \in M\}$$

$$Z_t = [d(X(i), X(j))]^2 - [d(X(k), X(\ell))]^2 = d_{ij} - d_{k\ell}$$

Pour tout t tel que $k(t) = (i,j,k,\ell)$

On se propose de construire l'application X de I dans V de telle sorte que soient vérifiées de façon "optimale" les contraintes d'isotonie

(a) $Z \geq 0$

qui sont équivalentes à (1).

La projection de tout vecteur de \mathbb{R}^m sur l'orthant positif $\{Z ; Z \in \mathbb{R}^m ; Z \geq 0\}$ s'écrit de façon classique \hat{Z} de composante t :

$\tilde{Z}_t = \text{Max}(0, Z_t)$. On est donc amené à résoudre un problème de regression sous contrainte de positivité dans \mathbb{R}^m .

De LEEUW propose de minimiser la fonction d'écart

$$\begin{aligned} F(q) &= \|Z - \tilde{Z}\|_q^q = \sum\{|Z_t - \tilde{Z}_t|^q ; t \in M\} \quad (q > 0, q \in \mathbb{R}) \\ &= \sum\{|Z_t - |Z_t||^q ; t \in M\} / 2^q \end{aligned}$$

qui s'annule si et seulement si Z appartient à l'orthant positif. L'invariance par homothétie est assurée en minimisant

$$L(q) = \frac{\|Z - \tilde{Z}\|_q^q}{\|Z\|_q^q} \quad \text{où } \|\cdot\|_q \text{ est la norme } L_q \text{ sur } \mathbb{R}^m$$

Remarque

Lorsque $q = 2$ l'expression de $L(2)$ est analogue à la forme de STRESSFORM1 initialement proposée par KRUSKAL [58,a], cependant on travaille dans \mathbb{R}^m et non \mathbb{R}^p .

L'expression de $L(2)$ est analogue à l'expression du rapport de corrélation $1 - r^2$, cependant la variable Z n'est pas centrée.

Interprétations de $L(q)$

Si l'on note :

$$M_+ = \{t ; t \in M, Z_t > 0\} \quad |M_+| = m_+$$

$$M_- = \{t ; t \in M, Z_t < 0\} \quad |M_-| = m_-$$

On a la relation

$$L(q) = \frac{\sum\{|Z_t|^q ; t \in M_-\}}{\sum\{|Z_t|^q ; t \in M\}} = \frac{\sum\{|Z_t|^q ; t \in M_-\}}{\sum\{|Z_t|^q ; t \in M_+\} + \sum\{|Z_t|^q ; t \in M_-\}}$$

Par suite lorsque q tend vers zéro on a

$$\varphi = \lim_{q \rightarrow 0} L(q) = \frac{m_-}{m_+ + m_-}$$

et donc de LEEUW développe une théorie de l'erreur qui bien que métrique tend à être ordinale lorsque q tend vers zéro. La minimisation directe de φ n'est pas aisée puisque c'est une fonction en escalier que l'on ne peut donc minimiser par des méthodes de Gradient. L'approche de de LEEUW est éclairée en posant

$$Y_t = \frac{1}{2} (1 - \text{signe}(Z_t)) \quad t \in M$$

qui est donc une indicatrice d'erreur, et par suite

$$\begin{aligned} L(q) &= \frac{\sum \{|Z_t|^q Y_t ; t \in M\}}{\|Z\|_q^q} \\ &= \sum_{t \in M} \frac{|Z_t|^q}{\|Z\|_q^q} Y_t \end{aligned}$$

l'erreur Y_t est donc pondérée par l'importance numérique de l'écart observé et de plus

i) $(0 \leq Y_t \leq 1) \Rightarrow (0 \leq L(q) \leq 1)$

ii) $(L(q) = 1) \Leftrightarrow (M = M_-)$

iii) $(L(q) = 0) \Leftrightarrow (M = M_+)$

iv) $(L(q) = \frac{1}{2}) \Leftrightarrow (F^+(q) = F^-(q))$

où

$$F^+(q) = \sum \{|Z_t|^q ; t \in M_+\} \quad F^-(q) = \sum \{|Z_t|^q ; t \in M_-\}$$

L'interprétation statistique de φ s'exprime en termes de corrélation de rang puisque le coefficient γ de KRUSKAL (William) et GOODMAN est défini par :

$$\gamma = \frac{m_+ - m_-}{m_+ + m_-} = 1 - 2\varphi$$

et s'identifie au τ de KENDALL lorsque Q est une ordonnance [8]. Par suite la valeur de φ_{Min} associée à la configuration optimale X^* peut être testée puisque la distribution de γ est connue [8].

de LEEUW se propose de minimiser $L(q)$ pour $q < 1$ car alors cette fonction est différentiable si $q > 0$ et le problème est résolu par des méthodes de projection (Gradient). L'approximation de φ est donnée par

$$|\varphi - L(q)| \leq \frac{1}{2} \sum_{t=1}^m \left| \frac{|z_t|^q}{\|Z\|_q^q} - \frac{1}{m} \right| = \frac{1}{2} \frac{\frac{1}{m} \sum_{t=1}^m \|z_t\|^q - \frac{1}{m} \sum_{t=1}^m |z_t|^q}{\|Z\|_q^q}$$

Cette relation obtenue par de LEEUW [33,c] en interprétant $L(q) = \sum_{t \in M} \left[\frac{z_t}{\|Z\|_q} \right]^q Y_t$ comme un contraste des Y_t permet d'écrire

$$(b) \quad |\varphi - L(q)| \leq \frac{1}{2} \cdot \frac{E_q(Z)}{V_q(Z)}$$

où $E_q(Z)$ est la moyenne arithmétique : $E_q(Z) = \frac{1}{m} \sum \{|z_t|^q ; t \in M\}$

$V_q(Z)$ est la variation totale : $V_q(Z) = \frac{1}{m} \sum \{|z_t|^q - E_q(Z)| ; t \in M\}$

La qualité de l'approximation de φ par $L(q)$ est donc contrôlable de façon simple et (b) montre qu'une condition nécessaire d'égalité tous les $L(q)$ lorsque q varie est l'égalité des $|z_t|^q$ non nuls.

Cette relation permet de plus de juger si la valeur de $L(q)$ reflète la taille ou le nombre des erreurs (auquel cas G_q est faible) grâce au gain :

$$G_q = \frac{\varphi}{1-\varphi} \cdot \frac{1-L(q)}{L(q)} = \frac{F^+(q)}{m_+} \cdot \frac{m_-}{F^-(q)}$$

Algorithme de minimisation de $L(q)$

Pour toute valeur de q différente de 1 $L(q)$ est différentiable en Z_t et l'on peut donc construire l'application X par une méthode itérative (Gradient).

$$X^{(s+1)}(i) = X^{(s)}(i) + (\alpha) \cdot \left[\frac{dL(q)}{dX} \right]^{(s)}(i)$$

où

$$\left[\frac{dL(q)}{dX} \right]^{(s)}(i) = \sum_{t \in M} \frac{\partial L(q)}{\partial Z_t} \cdot \frac{\partial Z_t}{\partial X(i)}$$

$$\frac{\partial L(q)}{\partial Z_t} = \frac{q}{\|Z\|_q} \{ [1 - 2L(q)] \text{Sign}(Z_t) - 1 \} |Z_t|^{q-1}$$

et $\frac{\partial Z_t}{\partial X(i)}$ est le vecteur de K-ième composante.

$$\begin{aligned} \left[\frac{\partial Z_t}{\partial X(i)} \right]_k = r \{ & (\delta_i^j - \delta_i^\ell) \text{signe}(x_k(j) - x_k(\ell)) |x_k(j) - x_k(\ell)|^{r-1} \\ & - (\delta_i^m - \delta_i^n) \text{signe}(x_k(m) - x_k(n)) |x_k(m) - x_k(n)|^{r-1} \} \end{aligned}$$

où $k(t) = (j, \ell, m, n)$ et δ_i^m est la fonction de Kroenecker.

Il résulte de cette expression que l'algorithme de de LEEUW est beaucoup plus lent que ceux exposés plus tôt à cause du temps de calcul de $\left[\frac{dL(q)}{dX} \right]^{(s)}(i)$. Les essais numériques d'un tel algorithme sont trop peu nombreux, à notre connaissance, pour juger de son comportement sur des données entachées de relations ultramétriques mais de par la définition de M il construit une représentation forte lorsqu'elle existe.

§ V - CONSTRUCTION D'UNE CONFIGURATION INITIALE

Les algorithmes décrits dans les deux paragraphes ci-dessus adoptent tous une stratégie itérative qui nécessite donc une initialisation. Leur sensibilité aux problèmes de minima locaux et de dégénérescence impose un choix rigoureux d'une configuration initiale, afin de limiter ces deux risques.

Deux buts nous semblent poursuivis par les algorithmes proposés. Outre la construction d'une configuration initiale $X^{(0)}$ on recherche de plus une estimation de r.

Notre présentation est construite suivant trois types d'algorithmes. Suivant le caractère probabiliste ou déterministe de leur fondement et suivant

que l'on travaille dans \mathbb{R}^p ou dans \mathbb{R}^m .

III-5-1 - Approches probabilistes

La proposition de KRUSKAL, de génération de $X^{(0)}$ par tirage de nombres au hasard, ne nous semble intéressante que pour tester l'"unicité" d'une solution associée à une valeur nulle de Ψ par replications de l'algorithme à partir de configurations initiales distinctes. Lorsqu'un minimum local est suspecté une telle démarche est valide.

Plus efficace nous semble être l'approche de BENZECRI [39] qui suppose que $X^0 = \{X^0(i), i \in I\}$ est un n échantillon issu d'un vecteur aléatoire de loi gaussienne sphérique $N(0, \sigma^2 I_r)$. Cette démarche présente l'inconvénient de supposer r connue. Alors d_b^2 suit une loi de khi-deux à r degrés de liberté et l'on montre que ces distances peuvent être reconstituées approximativement, à une constante multiplicative près à l'aide des inégalités induites par Q

$$d_b^2 = F_r^{-1} \left(\frac{1}{p} (R(b) - \frac{1}{2}) \right)$$

où pour tout b dans B

- i) $R(b)$ est la valeur de l'indice de rang $1 \leq R(b) \leq p = n(n-1)/2$
- ii) F_r^{-1} est la fonction inverse de la fonction de répartition associée à une loi du khi-deux à r degrés de liberté.

Utilisée dans le même contexte que l'approche de KRUSKAL, cette méthode nous semble de beaucoup préférable car elle construit une solution $X^{(0)}$ admissible sinon optimale. Elle est donc intéressante pour tester l'existence d'un minimum local. Son utilisation à posteriori est due au fait qu'elle nécessite de connaître r .

III-5-2 - Approches déterministes

La proposition initiale de KRUSKAL consistant à construire $X^{(0)}$ en prenant les n premiers termes de la suite $(1,0,\dots,0)$, $(0,1,\dots,0)$... $(0,\dots,0,1)$ $(2,0,\dots,0)$, $(0,2,\dots,0)$ etc... de vecteurs dans \mathbb{R}^r fut très longtemps un facteur d'échecs pour son algorithme et l'on préférera construire $X^{(0)}$ par des méthodes métriques qui permettent de donner une estimation de r et permettent de juger du gain apporté par l'affaiblissement des hypothèses qui fonde les méthodes ordinales.

Δ est alors supposé d'échelle d'intervalle et l'on ajuste un critère de "constante additive" conformément à ce que nous avons dit au chapitre I.

$$d_{ij} = T_{ij} + c \text{ où } T \text{ est un indice de distance isotone par rapport à } \Delta .$$

Si T est une matrice définie positive, on construit $X^{(0)}$ par analyse en composantes principales de la matrice de pseudo-produits scalaires

$$B = - \frac{1}{2} \left(I - \frac{E}{n} \right) T \left(I - \frac{E}{n} \right)$$

obtenue par double centrage de T, suivant en cela TORGERSON [17].

Lorsque tel n'est pas le cas la matrice de produits scalaires associés à la matrice D des distances carrés est

$$\begin{aligned} \Sigma &= - \frac{1}{2} \left(I - \frac{E}{n} \right) D^2 \left(I - \frac{E}{n} \right) \\ &= - \frac{1}{2} \left(I - \frac{E}{n} \right) (T + cI)^2 \left(I - \frac{E}{n} \right) \end{aligned}$$

où nous notons $(T+cI)^2$ la matrice de terme général $(T_{ij}+c)^2$

$$(b) \quad \Sigma = - \frac{1}{2} c^2 \left(I - \frac{E}{n} \right) \left(I - \frac{E}{n} \right) + 2 c B + M$$

où

$$M = - \frac{1}{2} \left(I - \frac{E}{n} \right) T^2 \left(I - \frac{E}{n} \right) \text{ et } T^2 = (T_{ij}^2)_{\substack{i=1,2,\dots,n \\ j=1,2,\dots,n}}$$

et $X^{(0)}$ est construite par analyse en composantes principales de Σ . On ne retient que les r dimensions associées aux r plus grandes valeurs propres de Σ . Les vecteurs propres associés vérifient :

$$t_{H_i H_i} = 1$$

$$t_{H_i \mathbf{1}} = 0 \quad \text{puisque } \Sigma \text{ est doublement centrée}$$

$$t_{H_i \Sigma H_i} = \lambda_i \quad \forall i \leq r$$

Dans une telle méthode on suppose toutes les valeurs propres positives ou nulles, c'est-à-dire que D est supposée être une distance euclidienne et l'on suppose qu'il existe r valeurs propres non nulles. On a alors :

$$(c) \quad \sum_{i=1}^n \Sigma_{ii} = \text{Trace}(\Sigma) = \sum_{i=1}^r \lambda_i = -\frac{r}{2} c^2 + 2c \sum_{i=1}^r t_{H_i B H_i} + \sum_{i=1}^r t_{H_i M H_i}$$

MESSICK et ABELSON [64] proposent alors l'algorithme

(1) Calcul de $C^{(0)}$ valeur initiale de c par

$$C^{(0)} = \text{Min} \{T_{ik} + T_{kj} - T_{ij} ; \quad i, j, k \in I \quad i \neq j\}$$

qui assure $T_{ij} - (T_{ik} + T_{kj}) - c \leq 0$

$$\Leftrightarrow d_{ij} - (d_{ik} + d_{kj}) \leq 0$$

(2) Calcul de $\Sigma^{(s)}$ par (b) et calcul de ses valeurs et vecteurs propres $\lambda_i \geq \epsilon$, ϵ nombre réel positif fixé

(3) Calcul d'une nouvelle valeur de c par (c) et aller en (2).

On s'arrête quand le processus se stabilise.

Une telle méthode méconnaît la possibilité de valeurs propres négatives assurant que D n'est pas une distance euclidienne, et peut donc conduire à une piètre configuration initiale. Aussi COOPER [45] propose une solution aux moindres carrés en ajustant le modèle

$$T_{ij} = d_{ij} + c + \epsilon_{ij} \quad \text{où } \epsilon_{ij} \text{ est un terme d'erreur}$$

et l'on minimise

$$G = \sum_{i < j} \epsilon_{ij}^2$$

et donc

$$0 = \frac{\partial G}{\partial c} = - \sum_{i < j} [T_{ij} - d_{ij} - c] \Leftrightarrow \hat{c} = \frac{1}{p} \sum_{i < j} (T_{ij} - d_{ij})$$

et puisque l'on peut toujours imposer $\sum_{i < j} T_{ij} = 0$ on prend $\hat{c} = \frac{1}{p} \sum_{i < j} d_{ij}$

La méthode de Gradient utilisée pour construire $X^{(0)}$ par minimisation de G ne permet cependant pas de connaître r. Aussi COOPER a t'il recourt à une analyse en composantes principales de $X^{(0)t} X^{(0)}$ pour déterminer r.

Le théorème de GUTTMANN (III.7) nous a en fait conduit à résoudre le problème de la constante additive en prenant pour T l'indice de rang. L'analyse en composantes principales de la matrice M_{λ_n} nous permet donc de donner une solution plus satisfaisante que celle de COOPER par un moindre coût algorithmique et l'expression explicite de la constante additive: $c = -2\lambda_n$ où λ_n est la plus petite valeur propre non nulle de la matrice de pseudo produits scalaires associés à la matrice des indices de rangs.

Cette méthode très liée à l'approche de GUTTMAN ([34] [52]) est construite par une argumentation théorique et non par analogie. Aussi nous semble t'elle préférable à la démarche de celui-ci.

Cependant le calcul explicite de λ_n nous semble important à un autre point de vue. L'existence de grandes valeurs propres négatives doit nous conduire à remettre en cause le choix d'une distance euclidienne pour analyser les données. On posera alors le problème du choix d'une métrique L_p avec

$p \neq 2$. [C'est ainsi qu'opère IHM (1970) dans "Distances et similitude en taximétrie" in Archéologie et Calculateurs Paris CNRS].

III-5-3 - L'approche de de LEEUW

Ceci se particularise par l'espace dans lequel on travaille. On considère deux matrices dans $\mathcal{M}_{m \times n}(\mathbb{R})$ définies par :

$$S_{bq} = (\delta_q^i - \delta_q^j) + (-\delta_q^k + \delta_q^l) = S_{bq}^1 - S_{bq}^2$$

$$\Sigma_{bq} = (\delta_q^i - \delta_q^j) + (\delta_q^k - \delta_q^l) = S_{bq}^1 + S_{bq}^2$$

où b est l'élément de M défini par $k(b) = (i, j, k, l)$ et δ_q^l est la fonction de Kronecker. Alors pour tout vecteur X dans \mathbb{R}^n on a :

$$G(X) = {}^t_X {}^t_{\Sigma S} X = \sum_{(i,j,k,l) \in M} (X(i) - X(j))^2 - (X(k) - X(l))^2$$

et donc minimiser $F(2)$ équivaut à minimiser $G(X)$ et de plus

$${}^t_{\Sigma S} = \frac{1}{2} [{}^t_{\Sigma S} + {}^t_{S \Sigma}] + \frac{1}{2} [{}^t_{\Sigma S} - {}^t_{S \Sigma}]$$

où

${}^t_{\Sigma S} - {}^t_{S \Sigma}$ est une matrice antisymétrique et par suite

$${}^t_X {}^t_{\Sigma S} X = \frac{1}{2} {}^t_X [{}^t_{\Sigma S} + {}^t_{S \Sigma}] X$$

On minimise $G(X)$ sous la contrainte ${}^t_{XX} = 1$ en minimisant le Lagrangien

$$L(X) = {}^t_X {}^t_{\Sigma S} X - \lambda [{}^t_{XX} - 1]$$

Le minimum est atteint lorsque

$$0 = \frac{\partial L(X)}{\partial X} = {}^t_{\Sigma S} X - \lambda X = [{}^t_{\Sigma S} + {}^t_{S \Sigma}] X - 2\lambda X = 2 [QX - \lambda X]$$

dont la solution est obtenue par calcul des valeurs propres.

Les matrices tS_1S_1 et tS_2S_2 sont doublement centrées, il en est de même de

$$Q = \frac{1}{2} [{}^t\Sigma S + {}^tS \Sigma]$$

Puisque $S = S_1 - S_2$

$$\Sigma = S_1 + S_2$$

Q est donc singulière et de valeurs propres

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k = 0 \geq \dots \geq \lambda_n$$

Le vecteur propre associé à λ_k est le vecteur constant 1 et tout vecteur propre associé à une valeur propre non nulle λ_i vérifie donc ${}^tX(i) \mathbb{1} = 0$. On peut donc écrire en revenant à $G(X)$ au minimum

$$G_r(X) = \sum_{k=1}^r \lambda_k = \sum_{(i,j,k,\ell) \in M} \left\{ \sum_{s=1}^r [x_s(i) - x_s(j)]^2 - [x_s(k) - x_s(\ell)]^2 \right\}$$

De plus puisque S_1 et S_2 sont par construction des matrices dont chaque ligne a seulement deux termes non nuls (un +1 et un -1) on a :

$$\text{Trace} ({}^tS_1S_1) = \text{Trace} ({}^tS_2S_2) = 2m$$

et par suite

$$\sum_{i=1}^n \lambda_i = \text{Trace} (Q) = 0$$

On prend donc le nombre de valeurs propres $\lambda_i \geq \epsilon$ comme estimation de r ($\epsilon \geq 0$).

§ 6 - L'ANALYSE a-DIMENSIONNELLE DES DONNEES DE PROXIMITE

Le développement de méthodes d'analyse des proximités se heurte à trois problèmes centraux :

- i) Détermination de la dimension r de l'espace vectoriel normé V
- ii) Détermination d'une distance d -définie sur V
- iii) Détermination d'une représentation optimale X de I dans V .

et le choix fait pour chacun de ces problèmes influence la solution réalisable des deux autres. Un choix arbitraire résoud le plus souvent les deux premiers points. L'attention s'étant jusqu'ici portée sur le troisième point.

Nous avons vu au paragraphe précédent que quelques informations peuvent être obtenues sur les deux premiers points lors de la construction d'une configuration initiale. Celle-ci fait le plus souvent intervenir des méthodes métriques et l'évaluation de r est fondée sur l'idée selon laquelle les méthodes métriques nécessitent en général une plus grande dimension r de V pour représenter les données sans perte d'information. Aussi ayant ainsi estimé r , on réalise plusieurs essais pour des valeurs voisines de r et l'on choisit la dimension minimisant l'indice de regression isotone.

De même le choix de la métrique L_p est remise en cause lors de la construction de la configuration initiale. Sa détermination est alors le plus souvent affaire de tâtonnements en termes de minimisation de l'indice de regression. Pourtant dans les deux cas les études par simulation ([69], [71] [73]) ont montré la faiblesse d'une telle démarche car la valeur de l'indice de regression dépend de r et d'autre part une bonne estimation de L_p nécessite d'être assuré au départ d'une bonne estimation de r et de la qualité de X^* (absence de minimum local ou de dégénérescence).

Pourtant nul n'a remis en cause le fondement du problème iii). Pourtant le théorème de HOLLMAN montre que le problème fondamental est l'approximation d'un indice de dissimilarité par une distance (à priori dans un espace de FINSLER de même que l'analyse hiérarchique pose le problème de l'approximation d'un indice de dissimilarité par une distance ultramétrique. La représentation par une carte (resp. un arbre) étant une conséquence de ce problème et non le problème fondamental.

C'est pourquoi nous proposons la démarche suivante :

- (a) Construction de la distance D meilleure approximation de Δ
- (b) Construction de la meilleure représentation X de I en termes de respect de la structure métrique induite sur V par D.

Outre la clarté des différents problèmes abordés, cette démarche permet de connaître la part de variance résiduelle due à l'approximation de Δ par D (et donc de l'adéquation d'un modèle multidimensionnel), la part de l'approximation d'une distance de FINSLER par une distance euclidienne ou plus généralement une distance L_p . Ce point nous permettra de mieux connaître la structure géométrique de l'espace de description de I.

Le point (a) constitue ce que nous appelons l'analyse a-dimensionnelle des données de proximité.

Notations

Soit \mathcal{C} le sous-ensemble de $B \times B \times B$ défini par

$$\mathcal{C} = \{(i,j,k) ; (i,j,k) \in B \times B \times B \quad i \neq j \neq k\}$$

On a $\text{Card}(\mathcal{C}) = n(n-1)(n-2)/2 = c$

Soit ℓ l'injection canonique de \mathcal{C} dans $B \times B \times B$.

A tout t dans \mathcal{C} on associe le vecteur S_t de \mathbb{R}^p de q-ième composante

$$S_{tq} = \delta_q^i + \delta_q^j - \delta_q^k$$

où δ_q^i est le symbole de Kronecker et $\ell(t) = (i,j,k)$.

On a alors pour tout vecteur D dans \mathbb{R}^p

$$s_t = {}^t S_t D = d_i + d_j - d_k$$

Le problème se formalise en termes de regression isotone : sur l'espace probabilisé $(B, \mathcal{S}(B), \lambda)$ est défini l'élément aléatoire Δ à valeurs dans l'ensemble ordonné (H, \mathcal{R}) . On se propose de construire l'élément

D de $L^2(\mathcal{F}(B))$ à valeurs dans $(\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+})$ tel que :

- i) D est la meilleure approximation au sens des moindres carrés de Δ
- ii) D vérifie les axiomes d'une distance.

B étant de cardinal fini on identifie $L^2(\mathcal{F}(B))$ et \mathbb{R}_+^P . D vérifie l'inégalité triangulaire s'il appartient au convexe fermé de \mathbb{R}_+^P défini par :

$$\xi = \{D ; D \in \mathbb{R}_+^P ; \sum_t s_t D \geq 0 \quad \forall t, \quad t = 1, 2, \dots, c\}$$

Afin que soit totalement déterminée on se limitera aux éléments du convexe fermé dans \mathbb{R}_+^P

$$\mathcal{F} = \xi \cap \{D ; D \in \mathbb{R}_+^P ; \sum \{d_b ; b \in B\} = \bar{D}/p\}$$

où \bar{D} est fixé (égal à 1 par exemple.)

Le problème est alors résolu en maximisant l'indice de regression isotone $v^2_{D/\Delta}$ sous la contrainte $D \in \mathcal{F}$. Ceci équivaut à minimiser le Lagrangien :

$$L(D) = \frac{\|D - \hat{T}\|_p}{\|D - \bar{D}\|_p} + \frac{\lambda}{p} \sum (s_t^2 ; s_t \leq 0)$$

où $\hat{T} = E(D/\mathcal{B})$ \mathcal{B} étant le sous σ treillis engendré par Δ . La différentiabilité de v^2 par rapport à D assure la convergence d'un algorithme de gradient si $\lambda \rightarrow \infty$. Lorsque le nombre d'itérations s croît ([83]).

On a alors le schéma :

$$(*) \quad d_i^{(s+1)} = d_i^{(s)} + \alpha^{(s)} \left[\frac{dL(D)}{dD} \right]^{(s)}(i)$$

où le gradient s'écrit :

$$\frac{\partial L(D)}{\partial d_i} = L(D) \left[\frac{d_i - \hat{T}_i}{\|D - \hat{T}\|_p} - \frac{d_i - \bar{D}}{\|D - \bar{D}\|_p} \right] + \frac{2\lambda}{p} \sum \{s_t S_{t_i} ; s_t < 0\}$$

On utilise alors l'algorithme

- (0) Initialisation $D^{(0)}$ sera construit par la méthode de BENZECRI avec $r = n-1$.
- (1) Calcul de $\hat{T}^{(s)}$ à l'aide de l'algorithme de regression isotone de Kruskal
- (2) Calcul de $D^{(s+1)}$ grâce à *
- (3) Si $v_{D/T}^2 \geq$ valeur test STOP
sinon $s = s+1$ aller en (1).

SHEPARD et CUNNINGHAM [68,f] ont développé récemment un algorithme dont les buts sont identiques, bien qu'à partir de motivations autres que les nôtres sous le nom de "MAXIMUM VARIANCE NONDIMENSIONAL SCALING" fondé sur le principe de maximisation de la variance pour assurer une faible dimension r de l'espace métrique sur lequel est définie D . Il maximise de façon heuristique la fonction

$$F(D) = \frac{1}{p} \{ \|D - \hat{T}\|_p^2 - W_1 \|D - \bar{D}\|_p^2 - W_2 \sum \{s_t ; s_t < 0\} \}$$

sous les mêmes contraintes sur D . Le schéma d'optimisation s'écrit alors

$$d_i^{(s+1)} = d_i^{(s)} + \alpha^{(s)} \left[\frac{dF(D)}{dD} \right]^{(s)}(i)$$

avec

$$\frac{\partial F(D)}{\partial d_i} = \frac{2}{p} \{ (d_i - \bar{D}) - W_1 (\hat{T}_i - d_i) - \sum \{s_t S_{t_i} ; s_t < 0\} \}$$

et propose de choisir $\alpha^{(s)}$ de telle sorte que

$$F(D^{(s)} + \alpha^{(s)} \left[\frac{dF}{dD} \right]^{(s)}) = \underset{\alpha}{\text{Max}} F(D^{(s)} + \alpha \left[\frac{dF}{dD} \right]^{(s)}) .$$

L'algorithme de CUNNINGHAM est identique au notre à ceci près que $D^{(0)}$ est défini par

$$d_i^{(0)} = b \left(1 - \frac{\delta_i}{\bar{\delta}}\right) + \bar{D} \text{ avec } 0 < b \leq \frac{\bar{\delta}\bar{D}}{\text{Max}_{b \in B} \{\delta_b\} - \bar{\delta}}$$

où $\bar{\delta} = \frac{1}{p} \sum_{b \in B} \delta_b$.

Un choix de b petit assure alors que $D^{(0)}$ est l'élément de \mathcal{F} . Notre démarche semble préférable à celle de SHEPARD tant par le choix de la configuration initiale que par la formulation en termes de régression isotone. De ce point de vue le retour de SHEPARD à son algorithme originel marque un recul théorique que l'on paie par un choix arbitraire de la constante W_1 . Cependant une confrontation algorithmique peut seule déterminer la démarche optimale.

§ 7 - L'ANALYSE DES DONNEES DE PROXIMITE CONDITIONNELLES

L'analyse des données de proximité conditionnelles étudie la description statistique de $\bigcup_i (B_i \times B_i)$ données agrégées. On dispose d'un tableau de données expérimentales :

$$\Delta : I \times I \rightarrow (H, \mathcal{R})$$

et l'on retient la relation de préordre Q_i induite par le profil :

$$\delta_i : \{i\} \times I \rightarrow (H, \mathcal{R})$$

sur l'ensemble

$$B_i = \{b_{ij} ; b_i \in B \quad h(b_i) = (i, j) \quad j \in I\}$$

On se propose de construire une application X de I dans un espace euclidien $V = (\mathbb{R}^r, d)$ de telle sorte que

- i) La dimension r de V soit minimale
- ii) Soient vérifiées de façon "optimale" les contraintes d'isotonie :

$$(1) \quad \sigma_{Q_i}(b_i, b'_i) (d_{b_i} - d_{b'_i}) \geq 0 \text{ Pour tous } (b_i, b'_i) \text{ dans } B_i \times B_i, \\ \text{pour tout } i \text{ dans } I$$

où σ_{Q_i} est la signature de la relation de préordre Q_i définie sur B_i .

III-14 - Définition

On appelle signature d'une relation de préordre Q définie sur un ensemble B , la fonction σ_Q définie sur $B \times B$, à valeurs dans $\{-1, 0, 1\}$:

$$\sigma_Q(b_1, b_2) = \begin{cases} +1 & \text{Si } Q[b_1, b_2] \text{ et } \bar{Q}(b_2, b_1) \\ -1 & \text{Si } \bar{Q}[b_1, b_2] \text{ et } Q(b_2, b_1) \\ 0 & \text{si } Q[b_1, b_2] \text{ et } Q(b_2, b_1) \\ 0 & \text{si } b_1 \text{ et } b_2 \text{ ne sont pas } Q\text{-comparables} \\ 0 & \text{si } b_1 = b_2 \end{cases}$$

où \bar{Q} est la relation non- Q

La préordonnance Q induite par $\{Q_i, i \in I\}$ sur I est une relation partielle et l'on cherche donc, par une argumentation de regression isotone, l'élément D de $\mathbb{R}_+^P \equiv L^2(\mathcal{P}(B))$ qui factorise à travers Δ par

$$d_{b_i} = m_i(\delta_{b_i}) \quad (\forall b_i, \quad b_i \in B_i), \quad (\forall i, \quad i \in I)$$

où $\{m_i, i \in I\}$ est une famille de fonctions isotones de (H, \mathcal{Q}) dans (\mathbb{R}_+, \leq) .

La résolution de ce problème nécessite donc d'utiliser une méthode de regression isotone par morceaux et l'on utilise comme indice de regression isotone soit $v_{D/\Delta}^2$ comme GLEASON [50], soit comme ROSKAM [14] :

$$L(X) = \left[\sum_{i \in I^*} m_i L_i(x) \right]^{1/2}$$

où m_i est la fréquence relative d'occurrence dans I du préordre Q_i et I^* est le sous-ensemble de I des éléments i associés à des préordres Q_i distincts

$$L_i(x) = \frac{\sum \{ [d_{b_i} - T_{b_i}]^2 ; b_i \in B_i \}}{\sum \{ [d_{b_i} - \bar{d}_i]^2 ; b_i \in B_i \}}$$

$L_i(X)$ est donc associée à l'indice de regression $v_{D/\Delta}^2$ dans B_i

$$L_i(X) = 1 - [v_{D/\Delta}^2]_{B_i}$$

et nous avons noté

$$\bar{d}_i = \frac{1}{|B_i|} \sum \{ d_{b_i} ; b_i \in B_i \}$$

T_{b_i} = la valeur en b_i de l'espérance conditionnelle de D par rapport au sous σ -treillis de $\mathcal{F}(B_i)$ engendré par le profil δ_i .

Lorsque $T^{(s)}$ est connu, le calcul de $X^{(s+1)}$ se fait de façon classique par minimisation de $L(X)$ par une méthode de Gradient :

$$X^{(s+1)}(i) = X^{(s)}(i) + \alpha^{(s)} \left[\frac{dL}{dX} \right]^{(s)}(i)$$

Une telle démarche généralise l'approche de KRUSKAL au cas de données conditionnelles. De même une méthode de projection est développée par de LEEUW dans \mathbb{R}^m [33,c] en associant à $\{Q_i, i \in I\}$ une partition de $M : M = \cup \{M_i, i \in I\}$ et en définissant l'indice d'écart :

$$F(q) = \sum \{ m_i F_i(Q) ; i \in I^* \}$$

$$F_i(q) = \sum \{ \|Z_t\|_q^{-q} (|Z_t| - Z_t)^q ; t \in M_i \}.$$

L'utilisation pratique de telles méthodes s'avère délicate à cause de leur sensibilité accrue aux problèmes de minima locaux et de dégénérescence! Ceci est dû au faible nombre de contraintes ordinales imposées par

$\{Q_i, i \in I\}$.

La fréquence de telles données a motivé la recherche de méthodes permettant d'éviter ces problèmes. SIBSON [70] et KENDALL (in [24]) proposent une méthode de régularisation des données fondée sur la remarque suivante :

Deux caractères i et j sont proches l'un de l'autre si non seulement δ_{ij} est faible mais si les profils δ_i et δ_j se ressemblent. On construit donc à partir du tableau Δ le tableau Γ de terme générique

$$\Gamma_{ij} = \frac{p}{2} (1 - \gamma_{ij})$$

où $p = n(n-1)/2$ et γ_{ij} est l'indice de corrélation de rangs de GOODMAN et KRUSKAL des classements Q_i et Q_j .

L'itération de ce processus donne le calcul de $\Gamma^{(2)}$ construit de façon identique à partir de Γ , puis de $\Gamma^{(2)}$ etc... L'intérêt d'une telle démarche est d'une part de "briser" les exaequos et de régulariser les données tout en conservant la structure ordinale.

Bien qu'actuellement non fondée en théorie, cette démarche donne des résultats intéressants en pratique. Cependant, il est clair qu'une telle approche conserve les structures ultramétriques lorsque celle-ci existent dans les données (car on a une isotonie conditionnelle, et ne prémunie pas contre les risques de dégénérescence).

Lorsque les méthodes présentées échouent, il est possible de faire une analyse en composantes principales de la matrice $\Gamma^{(n)}$ pour un n donné, ou de la matrice des rangs généralisés R . Cependant cette démarche n'est guère plus fondée théoriquement que l'approche de SIBSON. De plus le faible nombre de contraintes ordinales imposées par un tel modèle nous semble montrer les limites de méthodes ordinales dans l'appréhension de structures complexes.

Remarque

L'interprétation d'une carte construite par analyse des proximités conditionnelles est rendue délicate par le type de regression isotone utilisé. On a en effet :

$$m_i(\delta_{ij}) = d_{ij} = d_{ji} = m_j(\delta_{ji})$$

Pour tous (i,j) dans $I \times I$ et donc l'interprétation des proximités doit être réalisée de façon non symétrique.

CHAPITRE IV

ANALYSE DES DONNEES DE PROXIMITE NON AGREGEES

Nous étudions dans ce chapitre le problème de représentation de données de proximité GLOBALES et NON-AGREGEES, dans un espace euclidien de faible dimension.

Dans un premier paragraphe, nous définissons le problème et montrons ses liens avec l'analyse des données de proximité globales et agrégées.

Le second paragraphe étudie les approches semi-métriques qui furent proposées. Notre étude est centrée sur les travaux de CARROLL [30,d].

Dans le dernier paragraphe, nous montrons qu'une approche purement ordinale de ce problème est possible et nous proposons un algorithme de résolution. Notre modèle est comparé au modèle semi-métrique de CARROLL.

§ 1 - LE PROBLEME ETUDIE EN ANALYSE DES PROXIMITES NON-AGREGEES

On s'intéresse à la description d'un ensemble fini de caractères, I . Nous avons vu qu'en analyse des proximités agrégées, le statisticien dispose d'un unique tableau :

$$\Delta(\omega) : I \times I \rightarrow (H, \mathcal{S}_I)$$

qui est le plus souvent issu d'une expérimentation unique ω , ($\omega \in \Omega$).

Lorsque $\Delta(\omega)$ est issu de l'agrégation de plusieurs mesures expérimentales $\Delta(\omega_j)$ associées à un m-échantillon $\{\omega_j, j \in J\}$ d'éléments de Ω , l'analyse des proximités non agrégées a pour but de mettre en évidence les différences individuelles entre éléments ω_j de l'échantillon, à partir des informations ordinales contenues dans les tableaux $\Delta(\omega_j)$.

Dans un but simplificateur, nous identifions dans la suite, le sous-ensemble $\{\omega_j, j \in J\}$ de Ω à l'ensemble d'indices J et notons Δ^j pour $\Delta(\omega_j)$. Soit Q_j la préordonnance induite par Δ^j sur I . Le problème peut être formulé ainsi :

Sur un ensemble de cardinal fini I , sont définies m préordonnances Q_j ($j \in J$). On se propose de construire un espace Euclidien $V = (\mathbb{R}^r, d)$ et deux applications :

i) $X : I \rightarrow V$

ii) $Y : J \rightarrow V$

de telle sorte que :

i) la dimension r de V soit minimale

ii) Soient vérifiées de façon "optimale" les contraintes d'isotonie :

(1) $\sigma_{Q_j}(b, b') [d_b - d_{b'}] \geq 0 \quad (\forall (b, b') ; (b, b') \in B \times B \quad \sigma_{Q_j}(b, b') \neq 0) \quad (\forall j, j \in J)$
avec

$$(h(b) = (i, k)) \Rightarrow (d_b = d(X(i), X(k))).$$

iii) Y traduise en termes de dualité, l'écart de chacune des préordonnances Q_j au consensus Q_X induit par la représentation X .

La démarche adoptée pour traiter de telles données est un enrichissement de l'analyse des proximités agrégées, étudiée au chapitre précédent. En effet elle est fondée sur la notion de consensus entre les éléments j de J .

Lorsque l'existence d'un tel consensus est montrée, par des méthodes d'inférence statistique non-paramétriques ([6], [8], [12], [32]), on construit le compromis X dans V . La construction de Y , permet ensuite de juger du contenu informationnel de celui-ci et de la qualité de la représentation de chacune des préordonnances Q_j dans le consensus Q_X . On peut alors construire une représentation euclidienne de chacune des préordonnances Q_j bien représentées dans Q_X .

Remarque

La suite de notre exposé suppose assurée l'existence d'un consensus.

L'analyse et la structuration de données multimodales ont suscité de nombreux travaux, particulièrement pour la description de séries chronologiques multidimensionnelles [31] (alors J est identifié au temps). Ces travaux, fondés sur des méthodes d'analyse factorielle ([75,a]) supposent les données d'échelle intervalle.

Notre étude est limitée aux démarches utilisables pour traiter des données de proximité non agrégées ordinales.

§ 2 - RESOLUTION SEMI METRIQUE DU PROBLEME

La première approche semi-métrique de $B \times B$ données non agrégées, due à TUCKER et MESSICK [75,b], est fondée sur la notion de point idéal introduite par COOMBS [5]. Elle consiste en une analyse en composantes principale de la matrice des corrélations de rangs associée aux éléments j de J . On recherche dans la configuration obtenue s'il existe des groupes. Dans ce cas on construit des représentants "idéaux" de chacun des groupes (en termes de barycentre) et la description statistique du tableau des proximités inter points idéaux est alors réalisée par les méthodes usuelles d'analyse des données. Elle permet ainsi de confronter les "points de vue" de chacun des groupes.

Cependant une telle méthode ne fait aucune hypothèse explicite d'existence d'un espace commun de représentation pour l'ensemble des éléments j dans J . Une telle hypothèse est le fondement des travaux ultérieurs ([30,a,b,d], [43], [56,a]) que nous étudions. Celles-ci furent étudiées dans le cadre de données d'intervalle et se proposent d'ajuster un modèle linéaire, au sens des moindres carrés, ce que l'on note :

$$(o) \quad D^{[2]}(j) \cong L_j \circ \Delta(j) \quad (\forall j, j \in J)$$

où $D^{[2]}(j)$ est la matrice symétrique d'ordre n :

$$D^{[2]}(j) = \{ [d_{i\ell}^j]^2 \quad ; \quad (i, \ell) \in I \times I \}$$

et l'application affine L_j vérifie le "critère de la constante additive"

$$[d_{i\ell}^j]^2 = \delta_{i\ell}^j + c^j \quad \text{où } \delta_{i\ell}^j \text{ est l'élément générique de } \Delta(j).$$

La détermination de la constante c^j se fait pour tout j par l'une des méthodes exposées en III.5.2.

La détermination de X et Y est obtenue en estimant les paramètres du modèle posé par HORAN [56,a]

$$(oo) \quad [d_{i\ell}^j]^2 = {}^t[X(i)-X(\ell)] W(j) [X(i)-X(\ell)]$$

où la matrice $W(j)$, symétrique semi définie positive d'ordre r , associée à chaque élément j dans J , une métrique euclidienne dans V (de dimension fixée r !). La résolution d'un tel modèle dans toute sa généralité est mise en oeuvre dans l'algorithme IDIOSCAL dû à CARROLL ([30,a,d], [43], [27]).

Cependant, la proposition III-8 montre qu'un modèle aussi général est inadapté à l'étude de données ordinales et suivant l'approche primitive de CARROLL ([30,b]) dans INDSCAL, nous supposons $W(j)$ diagonale et notons :

$$W(j) = \text{diag} (y_k(j))$$

Le vecteur $Y(j)$ traduit alors en termes d'homothétie le poids dans V de chacun des axes de représentation pour l'élément j de J .

IV-2-1 - Algorithme métrique

La matrice $\Sigma(j)$ de produits scalaires, associée à $D^{[2]}(j)$, est obtenue par double centrage (cf. III.5.2.)

$$\Sigma(j) = -\frac{1}{2} \left[I - \frac{E}{n} \right] D^{[2]}(j) \left[I - \frac{E}{n} \right]$$

d'élément générique :

$$(1) \quad \sigma_{i\ell}(j) = \sum_{k=1}^r y_k(j) x_k(i) x_k(\ell)$$

La détermination des paramètres Y et X se fait dans INDSCAL par décomposition du tableau trimodal :

$$\Sigma = \{ \Sigma(j) ; j \in J \}$$

défini sur $I \times I \times J$, grâce à la méthode de relaxation par blocs due à WOLD [78] .

Identifications

i) Σ est identifié à la matrice Σ^* à n^2 lignes et m colonnes de terme générique :

$$\sigma^*((i,\ell),j) = \sigma_{i\ell}(j)$$

ii) Σ est identifié à la matrice Σ^{**} à $n.m$ lignes et n colonnes de terme générique

$$\sigma^{**}((i,j),\ell) = \sigma_{i\ell}(j)$$

Notations

Soient Z (resp. W) la matrice à n^2 (resp. $n.m$) lignes et r colonnes de terme générique :

$$z_{(i,\ell),k} = x_k(i) \cdot x_k(\ell)$$

$$w_{k,(i,j)} = x_k(i) \cdot y_k(j)$$

l'équation (1) s'écrit de façon équivalente

$$(2) \quad \Sigma^* = ZY$$

$$(3) \quad \Sigma^{**} = WX$$

Y et X sont obtenus par ajustement au sens des moindres carrés des relations (2) et (3) suivant l'algorithme agissant à la s -ième itération comme suit :

(a) Calcul de $Y^{(s)}$ dans la relation (2)

$$Y^{(s)} = [t_Z^{(s-1)} \quad z^{(s-1)}]^{-1} t_Z^{(s-1)} \Sigma^*$$

(b) Calcul de ${}_1X^{(s)}$ dans la relation (3)

$${}_1X^{(s)} = [t_{2W}^{(s)} \quad {}_2W^{(s)}]^{-1} t_{2W}^{(s)} \Sigma^{**}$$

(c) Calcul de ${}_2X^{(s)}$ dans la relation (3)

$${}_2X^{(s)} = [t_{1W}^{(s)} \quad {}_1W^{(s)}]^{-1} t_{1W}^{(s)} \Sigma^{**}$$

où ${}_eW$ est définie de façon identique pour $e = 1$ et 2 par

$${}_e w_{k,(i,j)} = {}_e x_k(i) y_k(j)$$

distinguant ainsi dans (1) l'estimation de ${}_1x_k(i)$ lorsque $y_k(j)$ et ${}_2x_k(l)$ sont fixés de celle de ${}_2x_k(l)$ lorsque $y_k(j)$ et ${}_1x_k(i)$ sont fixés.

Remarques

- i) L'algorithme est interrompu lorsqu'une solution stable $(\hat{{}_1X}, \hat{{}_2X}, \hat{Y})$ est atteinte.
- ii) Rien n'impose l'égalité de ${}_1X^{(s)}$ et ${}_2X^{(s)}$ en cours d'itération. Cependant la symétrie des données :

$$\sigma_{il}(j) = \sigma_{li}(j)$$

assure qu'à l'optimum, il existe une matrice régulière diagonale C telle que

$$(*) \quad \hat{{}_1X} = C \hat{{}_2X}$$

et puisque $\hat{{}_2X}$ est estimée en dernier, on posera $\hat{X} = \hat{{}_2X}$.

De plus en imposant algorithmiquement (*) on corrigera le biais éventuel dans l'estimation \hat{Y} de Y.

Normalisations

- i) On assure la normalisation des données en imposant à Σ^* de vérifier

$$t_{\mathbb{1}} \mathbb{1}_{[\Sigma^*]}^{[2]} = t_{\mathbb{1}} \mathbb{1}_m \quad \text{où } \mathbb{1} \text{ est le vecteur de composantes égales à 1}$$

On assure ainsi l'égalité des variances et donne donc une égale importance dans l'analyse à tout j dans J.

- ii) Une normalisation des représentations est imposée par l'invariance par homothétie de (1), si l'on définit :

$$\tilde{y}_k(j) = \lambda^2 y_k(j)$$

$$\tilde{x}_k(j) = \lambda^{-1} x_k(j) \quad \lambda \neq 0$$

La normalisation de X assurant l'égalité des variances sur chacun des axes de V assure une interprétation aisée de la représentation Y. Le carré de la distance à l'origine de Y(j) pour tout j peut être "approximativement" interprété comme la part de variance du caractère j expliquée dans le compromis X. Le vocable "approximativement" est employé pour marquer que nous ne sommes pas assurés ici de la non corrélation des axes factoriels dans V. Cependant lorsque la normalisation iii) des données est assurée, la variance du caractère j est totalement expliquée si Y(j) est à distance unité de l'origine.

IV-3-2 - Algorithmique semi-métrique

L'analyse de données ordinales, proposée par CARROLL, consiste à itérer l'ajustement du modèle linéaire (0) à la factorisation isotone de $D^{[2]}(j)$:

$D^{[2]}(j) = m_j \circ \Delta(j)$ où m_j est une fonction isotone de (H, Q_j) dans (\mathbb{R}_+, \leq) .

A chaque pas s de l'itération, on construit le vecteur $T^{(s)}(j)$ dans \mathbb{R}_+^p , regression isotone de $D^{(s)}(j)$, suivant l'algorithme exposé en II.3. La construction de X et Y dans l'espace euclidien V, de dimension r fixée, est alors conduite par l'algorithme :

(o) Initialisation

Pour tout j dans J on calcule la matrice des rangs généralisés R(j) associée à $\Delta(j)$ et l'on pose :

$$T^{(0)}(j) = R(j)$$

(1) Phase métrique ($s \geq 1$)

(1-a) Calcul de $\Sigma^{(s)}$, à partir de $T^{(s-1)} = \{T^{(s-1)}(j) ; j \in J\}$
par calcul de la constante additive et double centrage .

(1-b) Calcul de $(X^{(s)}, Y^{(s)})$ d'après IV-2-2.

(1-c) Calcul de $(D^{[2]}(j))^{(s)}$ par (0,0) pour tous j de J.

(2) Phase ordinale

(2-a) Calcul de $T^{(s)}(j)$, regression isotone de $D^{[2]}(j)$ (cf. II.3), pour
tout j dans J.

(2-b) Calcul de l'indice de regression isotone $v^2(j)$ associé à chaque
élément j dans J.

(2-c) En déduire l'indice global

$$\psi^2 = \sum_{j \in J} [1 - v^2(j)]$$

si $\psi^2 \leq \epsilon$ fin de l'algorithme.

sinon aller en (1) en posant $s = s+1$.

Remarques

La proposition III-8 montre que l'introduction de la phase ordinale détruit le caractère unique du système d'axes de représentation.

Les deux phases de l'algorithme sont de nature très différentes puisque la première ajuste des produits scalaires alors que la seconde ajuste des distances et CARROLL conclue à l'impossibilité de mener une analyse purement ordinale par une telle démarche.

Cependant l'application qui à $D^{[2]}(j)$ associe $\Sigma(j)$ par double centrage est antitone et l'on peut envisager de construire, en phase (2), la regression antitone $\Xi(j)$ de $\Sigma(j)$.

L'indice de regression isotone est alors identique au STRAIN introduit par CARROLL [30,b] puisque $\Sigma(j)$ est doublement centrée.

$$1 - v^2_{\Sigma(j)/\Delta(j)} = \frac{\|\Sigma(j) - \Xi(j)\|_p^2}{\|\Sigma(j)\|_p^2}$$

La mise en oeuvre de l'analyse semi-métrique de CARROLL conduit en pratique à une représentation X (resp. Y) de I (resp. J) peu discernables de celle construite par des méthodes métriques. Ceci est dû tant à la robustesse de celles-ci (accrue par rapport au cas agrégé) aux non-linéarités mais aussi à la conception itérative de la démarche en deux phases.

L'algorithme est affaibli par la recherche conjointe de X et Y . Nous montrons qu'une présentation "naturelle" des algorithmes purement ordinaux, nécessite de construire X dans V , puis Y .

§ 3 - RESOLUTION PUREMENT ORDINALE DU PROBLEME

Seuls KRUSKAL et Mc GEE ([35],[63,b]) ont, à notre connaissance, abordé ce problème, dans un cas particulier puisque l'application Y n'est pas construite.

A la démarche triviale consistant à construire pour tout j dans J l'application ${}_jX$ dans un même espace euclidien V , par regression isotone par morceaux

$$D(j) = m_j \circ \Delta(j),$$

KURSKAL oppose un modèle interprétant $\{\Delta(j), j \in J\}$ comme un échantillon issu de Δ et associe à chaque élément j de J sa propre représentation ${}_jX$ par regression isotone :

$$D(j) = m \circ \Delta(j)$$

sans aucune contrainte de similarité des ${}_jX$ [in 30,b].

L'approche de Mc GEE a pour but de contrôler le degré de similarité des $_j X$ en utilisant comme paramètre, la somme des carrés des différences des coordonnées. Ce paramètre varie entre deux valeurs extrêmes correspondant aux deux cas étudiés par KRUSKAL. Il est cependant trop global pour exprimer comment deux configurations $_j X$ et $_k X$ diffèrent de l'identité.

Nous répondons à cette question en termes de dualité, en construisant un modèle analogue à INDSCAL mais purement ORDINAL, en deux phases :

- i) Construction de X dans V de façon à respecter les contraintes d'isotonie.
- ii) Construction de Y dans le dual de V sous contraintes d'isotonie.

IV-3-1 - Construction de V et de X

La formulation du problème en termes de regression isotone généralise celle de l'analyse des données de proximité agrégées au cas d'une famille de préordonnances.

On considère sur l'espace probabilisé $(B, \mathcal{P}(B), \lambda)$ (les notations étant celles du chapitre III) les éléments aléatoires suivants :

- i) Une famille finie de données $\{\Delta(j), j \in J\}$, $\Delta(j)$ étant à valeurs dans l'ensemble (H, \mathcal{R}) .
- ii) La distance D à valeurs dans l'espace probabilisable $(\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+})$.

Nous notons \mathcal{B} le sous σ -treillis de $\mathcal{P}(B)$ engendré par $\{\Delta(j), j \in J\}$ ou, de façon équivalente, par la famille finie de préordonnances $\{Q_j, j \in J\}$ définies sur I . X étant une représentation de I dans l'espace euclidien $V = \mathbb{R}^r$. On recherche l'élément D dans $D(I)$ (cf. Prop. III.9) et l'élément T de $L^2(\mathcal{B})$ qui réalisent le minimum de

$$\psi^2 = 1 - \nu^2 = \frac{E\{|D-T|^2\}}{\text{Var}(D)}$$

Lorsque D est connue, T est l'espérance de D conditionnellement à \mathcal{B} et deux approches sont envisageables pour construire D et donc X.

i) La première méthode consiste en une approche directe, fondée sur un algorithme de projection analogue à ceux étudiés au chapitre précédent.

a) Initialisation

Construction de $D^{(0)}$ d'après le III-5.

b) Au pas s ($s \geq 1$)

Calcul de $T^{(s)}$ espérance conditionnelle de $D^{(s-1)}$ sachant \mathcal{B} par l'algorithme de GEBHARDT (II.3.1.).

c) Calcul de $D^{(s)}$ par minimisation de ψ^2 (par la méthode de gradient étudiée en III.3.3.).

d) Calcul de ψ^2 et test d'arrêt.

Si $\psi^2 \leq \varepsilon$ STOP

Sinon $s = s+1$, aller en (b)

ii) La seconde méthode consiste en une approche indirecte fondée sur les résultats du paragraphe II.3.1. On construit ainsi un consensus Q à partir duquel on peut grâce à une méthode d'analyse des proximités agrégées, construire V et X.

Remarques

i) Cette démarche est analogue à l'approche métrique de HORAN [56,a] qui construit le compromis en remarquant que si (oo) est vrai pour tout j

$$D_o^{[2]} = \frac{1}{n} \sum_{j \in J} D^{[2]}(j)$$

définit une distance euclidienne sur V avec :

$$[d_{i\ell}^0]^2 = {}^t [X(i)-X(\ell)] W_0 [X(i)-X(\ell)]$$

où

$$W_0 = \frac{1}{m} \sum_{j \in J} W(j)$$

- ii) La seconde stratégie est recommandable en pratique, devant le coût élevé de mise en oeuvre de l'algorithme de GERHARDT. De plus elle permet de tester l'existence d'un consensus, grâce aux méthodes usuelles de statistique non paramétrique.

IV-3-2 - Construction de Y

L'équation (oo) s'écrit

$$\begin{aligned} [d_{i\ell}^j]^2 &= {}^t [X(i)-X(\ell)] W(j) [X(i)-X(\ell)] \\ &= \sum_{k=1}^r y_k(j) (x_k(i)-x_k(\ell))^2 \end{aligned}$$

Lorsque l'on suppose que W(j) est une matrice diagonale d'ordre r

$$W(j) = \text{diag} (y_k(j))$$

Pour tout b dans B vérifiant h(b) = (i, l) on pose

$$(\star\star) \quad z_k(b) = (x_k(i)-x_k(\ell))^2$$

Le vecteur Z(b) vérifie dans V la relation :

$$(ooo) \quad [d_b^j]^2 = {}^t Z(b) Y(j) \quad (\forall j, j \in J)$$

Par suite si le modèle (oo) est justifié, la détermination de Y(j) est obtenue, pour tout j dans J, par mise en oeuvre du modèle linéaire ordinal étudié au paragraphe II.4.

Sur l'espace probabilisé $(B, \mathcal{S}(B), \lambda)$ sont définies :

- i) La variable ordinale $\Delta(j)$ à valeurs dans l'ensemble ordonné (H, \mathcal{R}) .
- ii) Le vecteur aléatoire Z à valeurs dans $(\mathbb{R}^r, \mathcal{B}_{\mathbb{R}^r})$ de carré intégrable $Z \in L_{\mathbb{R}}^2 r(\mathcal{S}(B))$.

On se propose de construire l'élément $Y(j)$ du dual de $L_{\mathbb{R}}^2 r(\mathcal{S}(B))$ tel que

$$D^{[2]}(j) = \langle Z, Y(j) \rangle$$

factorise de façon isotone à travers $\Delta(j)$:

$$D^{[2]}(j) = m_j \circ \Delta(j)$$

où m_j est une fonction isotone sur (H, \mathcal{R}) .

Remarques

- i) Le modèle défini par (ooo) est un affinement du modèle de Mc GEE et l'équivalent ordinal du modèle métrique INDSCAL et CARROLL.
- ii) La relation antitone associant $\Sigma(j)$ à $D^{[2]}(j)$ permet d'étudier le modèle (ooo) où l'on définirait

$$z_k(b) = x_k(i) \cdot x_k(\ell)$$

Cependant le modèle posé semble plus fondé, en psychométrie, en termes de modèle compensatoire (<<compensatory model>>[5], [14]) avec $z_k(b)$ défini par (xx).

- iii) Le choix d'un tel modèle pose le problème de la signification du système d'axes définis dans V . Compte tenu de la proposition III.8, le statisticien est libre d'en faire un choix arbitraire par exemple en termes

- i) d'axes de plus grande dispersion (analyse en composantes principales).
- ii) d'axes associés aux classifications (varimax).

L'algorithme de projection dans \mathbb{R}^p , donné au paragraphe II-4, permet de construire totalement la solution pour chacun des éléments j dans J , par une méthode de gradient. Nous montrons qu'une formulation du problème, analogue à celle de l'orthant positif étudiée plus haut, permet de donner une solution par programmation linéaire, en adaptant à ce problème l'approche de SRINIVASAN et SHOCKER de l'estimation des paramètres du modèle du "critère composite" à partir de comparaisons par paires [72,b]. L'indice j est ignoré dans notre exposé.

Notations

Nous reprenons ici les notations du paragraphe III.4.2.
Le sous-ensemble M de $B \times B$ défini par :

$$M = \{(b,b') ; (b,b') \in B \times B, \sigma_0(b,b') \geq 0\}$$

est identifié à $\{1,2,\dots,m\}$ où $m = |M|$
et l'on note k l'injection canonique de M dans $B \times B$. Si l'on pose

$$U_t = d_b^2 - d_{b'}^2,$$

Pour tout t dans M vérifiant $k(t) = (b,b')$, on définit les deux sous-ensembles de M :

$$M_+ = \{t ; t \in M \quad U_t > 0\} \quad |M_+| = m_+$$

$$M_- = \{t ; t \in M \quad U_t < 0\} \quad |M_-| = m_-$$

Algorithme

L'algorithme de SRINIVASAN et SHOCKER [72,b] se propose, en fait, de minimiser la fonction :

$$L(1) = \frac{\sum\{|U_t| ; t \in M_-\}}{\sum\{|U_t| ; t \in M\}} = \frac{\sum\{|U_t| ; t \in M_-\}}{\sum\{|U_t| ; t \in M_+\} \sum\{|U_t| ; t \in M_-\}} = \frac{1}{1 + \frac{F^+(1)}{F^-(1)}}$$

où $F^+(1) = \sum\{|U_t| ; t \in M_+\}$

$F^-(1) = \sum\{|U_t| ; t \in M_-\}$

il est donc équivalent de minimiser (1) et $\frac{F^-(1)}{F^+(1)}$, but poursuivis par les auteurs.

De la relation

$$d_b^2 = \sum_{k=1}^r z_k^{(b)} y_k,$$

il résulte que :

- i) $Y = 0$ réalise le minimum de $F^-(1)$, bien que n'apportant aucune information.
- ii) $F^{-1}(1)$ est invariante par homothétie sur Y car

$$F^{-1}(1) = \sum\{|U_t| ; t \in M_-\} = \sum\{\text{Max}(0, -U_t) ; t \in M\}$$

Pour lutter contre ces deux difficultés on impose la contrainte technique :

(c1) $F^+(1) - F^-(1) = 1$

c'est-à-dire

$$\sum\{U_t ; t \in M\} = 1$$

Remarque

En toute généralité on devrait poser

$$F^+(1) - F^-(1) = h \quad h > 0$$

Cependant on montre ([72,a]) que la solution Y^* ne dépend pas de h , auquel on affecte la valeur conventionnelle 1.

La définition de U_t permet d'écrire le problème de minimisation de $\frac{F^-(1)}{F^+(1)}$ en termes de programmation linéaire :

$$U_t = d_b^2 - d_{b'}^2 = \sum_{k=1}^r (z_k(b) - z_k(b')) y_k = \sum_{k=1}^r w_k(t) y_k$$

En notant pour tout t vérifiant $k(t) = (b, b')$:

$$w_k(t) = z_k(b) - z_k(b')$$

Par suite en posant :

$$a_k = \sum \{w_k(t) ; t \in M\},$$

La contrainte (c1) s'écrit :

$$1 = \sum_{t \in M} U_t = \sum_{k=1}^r \{ \sum_{t \in M} w_k(t) \} y_k = \sum_{k=1}^r a_k y_k$$

Remarque

Cette relation montre le caractère technique de la contrainte (c1) car puisque Y n'est pas identiquement nul, il existe un k_0 pour lequel ou bien y_{k_0} et a_{k_0} sont non négatifs, ou bien y_{k_0} est quelconque et a_{k_0} est non nul et on posera alors

$$y_{k_0} = \frac{1}{a_{k_0}}$$

$$y_k = 0 \text{ pour tout } k, k \neq k_0$$

qui fournit une solution réalisable au problème (bien que non optimale à priori).

On a de plus :

$$F^-(1) = \sum_{t \in M} \text{Max}(0, -U_t) = \sum_{t \in M} \text{Max}(0, -\sum_{k=1}^r w_k(t) y_k)$$

et si l'on pose pour tout t dans M , b_t la variable d'écart telle que

$$\sum_{k=1}^r w_k(t) y_k + b_t \geq 0$$

On a la proposition démontrée dans [72,a]

IV-1 - Proposition

Le vecteur Y de V réalisant le minimum de $L(1)$ est solution du programme linéaire

$$(P) \quad \left\{ \begin{array}{l} \sum_{k=1}^r w_k(t) y_k + b_t \geq 0 \quad (\forall t, t \in M) \\ \sum_{k=1}^r a_k y_k = 1 \\ b_t \geq 0 \quad (\forall t, t \in M) \\ y_k \geq 0 \quad (\forall k, k \in \mathcal{K}, \mathcal{K} \subseteq \{1, 2, \dots, r\}) \\ \sum_{t \in M} b_t = V(\text{Min}) \end{array} \right.$$

Remarques

- i) La contrainte $y_k \geq 0$, pour tout k dans un sous-ensemble \mathcal{K} de $\{1, 2, \dots, r\}$, permet d'imposer, lorsque ceci est fondé concrètement, une structure particulière à la représentation Y de J dans \mathbb{R}^r .

ii) Par définition on a pour tout t dans M

$$b_t \geq \text{Max}(0, - \sum_{k=1}^r w_k(t) y_k)$$

On a donc égalité à l'optimum et par suite V et $F^{-1}(1)$ sont égaux, à leur optimum commun. De plus la contrainte (c-1) assure qu'est atteint l'optimum de $\frac{F^-(1)}{F^+(1)}$ et donc de $L(1)$.

iii) L'existence d'une solution finie avec $V^* \geq 0$ est assurée par le théorème 2 de SRINIVASAN et SHOCKER ([72,a] p. 367). Cependant en pratique la structure (P) montre que la résolution du problème dual rend l'algorithme plus performant.

iv) La valeur de b_t à l'optimum montre la qualité de la représentation Y^* construite puisque les b_t non nuls correspondent au non respect des conditions d'isotonie imposées par Q . Les exemples traités montrent que lorsque une intransitivité existe dans les données, elle apparaît en général par un b_t non nul à l'optimum.

L'orientation des axes de $V = (\mathbb{R}^r, d)$ obtenue par analyse du consensus est arbitraire et en général choisie afin de faciliter l'interprétation. Les pondérations optimales $Y(j)$ construites par l'approche de SRINIVASAN et SHOCKER dépendent de l'orientation choisie. Cependant les auteurs montrent ([72,a] § 3 (g), pp 354-6) que si l'on effectue une rotation des axes K alors :

$$Z(b) = Z(b)K \quad (\forall b, b \in B)$$

et la solution $Y(j)$ obtenue alors vérifie

$$Y(j) = Y(j) [{}^t K^{-1}]$$

et l'on a vu en III-4-2 que $L_j(1)$ est invariante par rotation.

Compte tenu de cette souplesse d'utilisation, du contrôle explicite possible de la nature de la solution $Y(j)$ obtenue, et de l'assurance d'une convergence en un nombre fini de pas de l'algorithme du simplexe, la démarche exposée nous semble préférable à l'approche développée en III.4.2 en termes de gradient. Le coût informatique de mise en oeuvre de l'un et l'autre conduisent à une même conclusion. Cependant il est à noter que $Y(j)$ est une solution optimale au sens de L^1 et non aux moindres carrés (L^2).

CHAPITRE V

ANALYSE DES DONNEES DE DOMINANCE

Nous étudions ici la description statistique de $U(A_j \times A_j)$ données de dominance, dans un espace euclidien de "faible" dimension.

La formulation mathématique du problème est donnée dans un premier paragraphe. Nous montrons qu'un tel formalisme recouvre plusieurs problèmes concrets, qu'il convient de distinguer pour choisir l'un des modèles interprétatifs exposés dans ce chapitre. Pour ce faire nous reprenons la classification de CARROLL et opposons l'analyse interne aux méthodes externes.

Les deux paragraphes suivants sont consacrés à l'analyse interne. Dans le premier, nous étudions les méthodes de représentation conjointe, en distinguant les généralisations du modèle de déploiement unidimensionnel, dû à COOMBS [5], des méthodes semi-métriques de projection du Permutoèdre et d'analyse des correspondances.

Le troisième paragraphe expose les applications du modèle linéaire ordinal, à l'analyse interne des données de dominance.

Au paragraphe suivant, consacré à l'analyse externe, nous montrons que l'on peut donner, en termes de régression isotone, un analogue purement ordinal aux approches métriques et semi-métriques proposées par SCHONEMANN [66] et CARROLL ([15],[30,c]).

§ 1 - LE PROBLEME ETUDIE EN ANALYSE DES DONNEES DE DOMINANCE

On s'intéresse à un ensemble fini d'objets I , sur lequel sont définies m relations de préordre $\{P_j, j \in J\}$, issues de données expérimentales. Deux cas sont rencontrés en pratique :

i) En analyse des "préférences individuelles" on dispose d'un tableau

$$\Delta: J \times I \rightarrow (H, \mathcal{Q})$$

et l'on retient la relation de préordre P_j induite par le profil (Def. I.9)

$$\delta_{j.} : \{j\} \times I \rightarrow (H, \mathcal{Q})$$

sur l'ensemble

$$A_j = \{a_j ; a_j \in \{j\} \times I\}$$

ii) En analyse des "données de comparaisons par paires" on dispose d'une famille de m tableaux :

$$\Delta^j : I \times I \rightarrow (H, \mathcal{Q})$$

qui induisent sur I une relation de préordre P_j sur l'ensemble A_j .

Remarques

i) Le fait de supposer la transitivité des comparaisons par paires, assure l'équivalence des deux types de données expérimentales. Cependant nous étudions dans la suite des algorithmes n'imposant pas cette restriction. Les intransitivités sont alors interprétées comme des erreurs.

ii) Dans notre exposé nous parlerons pour I d'ensemble d'OBJETS et pour J d'ensemble de VARIABLES. Cette nomenclature recouvre, en pratique, des problèmes de nature différente, tels que :

- (a) Plusieurs juges (éléments de J) classent un ensemble de stimuli (éléments de I).
- (b) Un individu classe un ensemble de stimuli (éléments de I) par rapport à différents attributs subjectifs (éléments de J).
- (c) Un ensemble de candidats (éléments de I) est classé suivant leurs performances respectives à une batterie de tests (éléments de J).
- (d) Un ensemble de joueurs de tennis (éléments de I) est classé suivant leurs scores durant les tournois (éléments de J) de l'année.

On se propose de construire un espace euclidien $V = (\mathbb{R}^n, d)$ et deux applications:

i) $X : I \rightarrow V$

ii) $Y : J \rightarrow V$

de telle sorte que

i) la dimension r de V soit minimale.

ii) Pour tout j dans J , soient vérifiées "de façon optimale" les contraintes d'isotonie :

$$(1) \quad \sigma_{P_j}(a_j, a'_j) [d_{a_j} - d_{a'_j}] \geq 0 \quad (\forall (a_j, a'_j), (a_j, a'_j) \in A_j \times A_j, \sigma_{P_j}(a_j, a'_j) \neq 0)$$

où σ_{P_j} est la signature de la relation de préordre P_j (Def. III.14).

La définition de la distance d_{a_j} ($a_j \in A_j$) dépend du modèle interprétatif choisi. Le modèle de représentation CONJOINTE construit l'image $X(i)$ (resp. $Y(j)$) de l'élément i dans I (resp. j dans J) en tant que points dans V . Alors on a :

$$d_{a_j} = d(X(i), Y(j)) \quad \text{si} \quad a_j = (j, i)$$

Ce modèle est celui utilisé pour la représentation "simultanée" en analyse des correspondances, par exemple. Le modèle de représentation FACTORIELLE construit, en termes de dualité, l'image $X(i)$ (resp. $Y(j)$) de l'élément i dans I (resp. j dans J) en tant que point (resp. direction vectorielle) dans V . Alors $d_{a,j}$ est la distance à l'origine de la projection de $X(i)$ sur la direction vectorielle de vecteur unité $Y(j)$. On retrouve ici le modèle de représentation "simultanée", utilisé en analyse des composantes principales, dans le cas de données d'échelle intervalle.

Le choix entre l'un ou l'autre modèle interprétatif est lié à la nature des données traitées. Les implications de ces deux modèles sont discutées dans le champ psychométrique, par ROSKAM [14] et COOMBS [5].

Notre présentation des méthodes d'analyse des données de dominance, suit la classification de CARROLL [30,a], suivant le degré d'information dont dispose le statisticien. L'analyse INTERNE ne suppose aucune information a priori pour construire conjointement X et Y . Si l'on connaît V et X , par une analyse antérieure, l'analyse EXTERNE des données permet de construire Y .

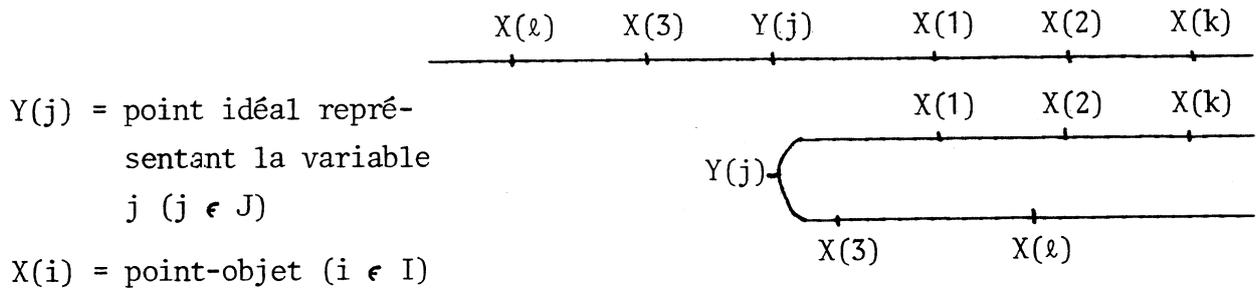
§ 2 - ANALYSE INTERNE : METHODES DE REPRESENTATION CONJOINTE

Cette approche est fondée sur la notion de point idéal, introduite par COOMBS [5], en analyse unidimensionnelle de données de dominance.

V.2.1 - Le modèle de déploiement unidimensionnel

COOMBS suppose qu'il existe un continuum sur lequel peuvent être représentés par un point tout objet et toute variable. Celle-ci est représentée sur le continuum par un POINT-IDEAL, objet "théorique" servant de référence pour classer les objets étudiés suivant la variable prise en compte. Les distances des points-objets à ce point idéal doivent alors respecter le classement associé à la variable.

Ceci revient à supposer l'existence d'un ordre BLACKTIEN ([1], [32]) obtenu par déploiement du classement associé à la variable étudiée.



Le positionnement des points X(i) et Y(j) sur ce continuum est obtenu, après testage statistique de l'existence d'un ordre BLACKTIEN [1], par une analyse algébrique connue sous le nom d'UNFOLDING (COOMBS [5] ch. 5 et 6) que les méthodes étudiées dans ce paragraphe tentent de généraliser au cas multidimensionnel.

V-2-2 - Le modèle de déploiement multidimensionnel

La première tentative de généralisation du modèle de COOMBS est due à BENNET et HAYS ([38], in [5] Ch. 7 et 12). Ceux-ci recherchent une représentation conjointe de I et J dans un espace euclidien de dimension minimale assurant le respect des contraintes d'isotonie (1). Leur démarche, algébrique, tente de déterminer les "axes de plus grande variabilité des classements". Elle n'est applicable que lorsque I est de cardinal petit et est non programmable. Cependant leur étude est la seule tentative d'estimation de la dimension r de V. Elle est fondée sur la

V-1 - Proposition

Si I est de cardinal n, la dimension r de l'espace de représentation V est supérieure ou égale à n-1 si et seulement s'il existe un ensemble de n! régions caractérisées par les n! permutations du groupe \mathfrak{S}_n .

Cette proposition, démontrée dans [3], permet de donner un critère simple d'estimation de r. Celle-ci est estimée par s-1 lorsque le plus grand groupe de permutations présent dans les données est \mathfrak{S}_s .

Remarques

- i) La notion de région, utilisée dans la proposition V-1, est définie comme le polyèdre convexe de V déterminé par les hyperplans médiateurs associés à tout couple de points dans V.
- ii) La démarche de BENNET suppose les données exactes. Lorsque celles-ci sont entachées d'erreurs, une telle méthode conduit à une estimation de r biaisée par excès.

Lorsque n est petit, on peut de plus juger de la qualité de l'estimation de r, en utilisant un autre résultat, dû à BENNET, (in [5] Ch. 12) et qui généralise la condition de BLACK au cas multidimensionnel.

V-2 - Proposition

Si V est r-dimensionnel tout sous-ensemble K de I de cardinal r+2 est décomposable en $K = K_1 \cup K_2$ de telle sorte qu'aucune variable j dans J ne classe tous les éléments de K_1 avant (resp. après) tous les éléments de K_2 .

La mise en oeuvre d'un tel critère nécessite cependant que n soit petit. Son coût devient, en effet, vite prohibitif.

La notion de point idéal permet à ROSKAM [14] et GLEASON [50] de donner une première mise en oeuvre algorithmique du modèle de déploiement multidimensionnel. En effet si l'on pose

$$I' = I \cup J$$

l'analyse d'un tableau de données de dominance Δ , s'interprète comme l'analyse en termes de données de proximité agrégées conditionnelles du tableau

$$\Delta' : I' \times I' \rightarrow (H, \mathcal{Q})$$

décomposable en :

$$\Delta' = \left[\begin{array}{ccc} \Delta'_1 & | & \Delta \\ \hline \Delta'_2 & | & \Delta'_3 \end{array} \right]$$

où Δ'_1 , Δ'_2 et Δ'_3 s'interprètent comme des données manquantes. Les résultats du paragraphe III.7 permettent alors de construire une représentation de I' et donc une représentation conjointe de I et J . Cependant l'existence de données manquantes accentue la sensibilité des méthodes étudiées en III.7, aux problèmes de dégénérescence. Aussi l'utilisation du modèle de l'orthant positif, applicable en toute généralité sur des $A \times A$ données, nous semble préférable (cf. III.4).

V-2-3 - Méthodes de projection

Les méthodes étudiées ici ne sont pas à proprement parler des méthodes ordinales, en ce sens que nous ne sommes pas assurés de la minimalité de la dimension de l'espace de représentation V . Cependant leur faible coût de mise en oeuvre et la richesse d'interprétation des configurations obtenues justifie leur emploi.

BENZECRI [3] propose d'utiliser l'analyse des correspondances pour le traitement de données de dominance, en étudiant conjointement les deux relations duales de préordre, induites par Δ . On dédouble alors l'ensemble I des objets en I^+ et I^- ensembles d'objets associés à chacune des relations ainsi obtenues. On note alors pour tout i^+ (resp. i^-) dans I^+ (resp. I^-) :

- i) $\frac{f_{ji^+}}{P_{j \cdot}}$ la fréquence absolue d'objets classés strictement avant i dans
- ii) $\frac{f_{ji^-}}{P_{j \cdot}}$ la fréquence absolue d'objets classés strictement après i dans

L'analyse du tableau de correspondances F ainsi obtenu ($F \in \mathcal{M}_{M \times 2n}(\mathbb{R}_+)$) permet d'obtenir une représentation conjointe de I^+ , I^- et J d'interprétation aisée. Une telle démarche nécessite cependant que m soit grand devant $2n$, afin d'assurer la stabilité des résultats.

Lorsque cette condition n'est pas vérifiée, on utilise une méthode de projection, fondée sur la notion de permutoèdre due à GUILBAUD et ROSENSTIEHL [18].

V-3 - Définition [1]

On appelle *DESACCORD* entre deux ordres totaux P_j et P_k sur I , tout couple (i, ℓ) d'éléments distincts de I tels que

ou bien $P_j(i, \ell)$ et $P_k(\ell, i)$

ou bien $P_j(\ell, i)$ et $P_k(i, \ell)$

Les deux relations d'ordre P_j et P_k sont *ADJACENTES* si elles présentent un seul désaccord.

V-4 - Définition [1]

On appelle *PERMUTOEDRE* de degré n l'ensemble des $n!$ ordres totaux sur un ensemble I de cardinal n , muni de la relation d'adjacence défini en V.3.

Remarque

Il résulte des deux définitions ci-dessus que les méthodes de représentation de données ordinales, fondées sur la notion de permutoèdre, sont limitées à l'étude de familles $\{P_j, j \in J\}$ de relations d'ordre.

De nombreuses présentations du permutoèdre ont été données ([18], [21] [32]). Il permet de donner dans \mathbb{R}^{n-1} une représentation conjointe de I et J . L'analyse en composantes principales du nuage ainsi construit, permet la construction d'une solution (V, X, Y) (BENZECRI [18], FENELON [21]). Nous donnons ici la construction d'une représentation (X°, Y°) de (I, J) dans

\mathbb{R}^{n-1} due à FENELON.

Dans \mathbb{R}^n muni de sa base canonique $\{e_i, i = 1, 2, \dots, n\}$, on pose

$$X^\circ(i) = e_i \quad (\forall i, i \in I)$$

On a alors la :

V-5 - Propriété

Pour tout i dans I , $X^\circ(i)$ appartient à l'intersection de la sphère de \mathbb{R}^n de rayon $[1 - \frac{1}{n}]^{1/2}$, centrée au point G de coordonnées $\frac{1}{n}$ et de l'hyperplan orthogonal en G au vecteur $\mathbf{1}_n$.

Démonstration

Le barycentre G des points $X^\circ(i)$, ($i \in I$) a pour coordonnées $\frac{1}{n}$ et dans le système d'axes obtenu par translation de l'origine en G , $X^\circ(i)$ s'écrit pour tout i dans I :

$$X^\circ(i) = e_i - \frac{1}{n} \mathbf{1}_n$$

où $\mathbf{1}_n$ est le vecteur de \mathbb{R}^n de terme générique égal à 1. Par suite

$${}^t \mathbf{1}_n X^\circ(i) = 0,$$

et $X^\circ(i)$ appartient à l'hyperplan de \mathbb{R}^n , passant par G et orthogonal au vecteur $\mathbf{1}_n$. De plus $X^\circ(i)$ a pour norme dans \mathbb{R}^n

$$\|X^\circ(i)\|_n^2 = \sum_{k=1}^n [x_k^\circ(i)]^2 = (-\frac{1}{n})^2(n-1) + (1 - \frac{1}{n})^2 = 1 - \frac{1}{n}$$

La représentation Y° de J dans \mathbb{R}^m est obtenue en associant au tableau Δ , le tableau des rangs R . La variable j est représentée dans le repère (G, \mathbb{R}^n) par le vecteur des rangs centrés, réduits, de k -ième composante :

$$y_k^\circ(j) = \left[\frac{n+1}{2} - r_{jk} \right] / \left[n(n-1)/n2 \right]^{1/2}$$

V-6 - Propriété

Pour tout j dans J , $Y^\circ(j)$ appartient à l'intersection de la sphère unité de \mathbb{R}^n centrée en G et de l'hyperplan orthogonal en G au vecteur $\mathbb{1}_n$.

Démonstration

$\frac{n+1}{2}$ est la moyenne empirique des $\{r_{jk}\}_{k=1,2,\dots,n}$. Par suite pour tout j dans J

$$t_{Y^\circ(j)} \mathbb{1}_n = 0$$

de plus $Y^\circ(j)$ étant le vecteur des rangs réduits, sa variance empirique $t_{Y^\circ(j)} Y^\circ(j)$ est égale à 1.

L'analyse en composantes principales du nuage de \mathbb{R}^{n-1} ainsi construit permet de donner une représentation conjointe de I et J dans un espace euclidien V de faible dimension. Son interprétation se fait en termes de rangs puisque l'on a la :

V-7 - Proposition

Le coefficient de corrélation de rangs entre deux variables j et j' est donnée sur le permutoèdre par

$$2(1-\rho(j, j')) = d_{jj'}^2 = d^2(Y^\circ(j), Y^\circ(j'))$$

Démonstration

Si j et j' sont deux éléments de J , on a sur le permutoèdre :

$$d_{jj'}^2 = \sum_{k=1}^n [y_k^\circ(j) - y_k^\circ(j')]^2 = \frac{12}{n(n^2-1)} \sum_{k=1}^n (r_{jk} - r_{j'k})^2 = 2(1-\rho(j, j'))$$

de plus dans le triangle défini par G, Y°(j) et Y°(j') on a :

$$d_{jj'}^2 = [2 \sin \frac{\alpha}{2}]^2 = 2(1 - \cos \alpha)$$

où α est l'angle des vecteurs $GY^\circ(j)$, $GY^\circ(j')$. Par suite le coefficient de corrélation de SPEARMAN $\rho(j, j')$ s'interprète comme le cosinus de l'angle formé par les deux variables j et j' sur le permutoèdre.

Remarque

JACQUET LAGREZE [32] montre que le coefficient de corrélation de KENDALL τ s'interprète sur le permutoèdre par :

$$\tau_{jj'} = 1 - 2 \frac{d_{jj'}}{\text{Max}(d_{jj'} ; (j, j') \in J \times J)} = \cos \theta_{jj'}$$

où $\theta_{jj'}$ est l'angle entre deux variables défini par GUILBAUD. Il en résulte que $\theta_{jj'}$ vérifie :

$$\theta_{jj'} = \frac{\pi}{2} (1 - \tau_{jj'})$$

et, comparée à l'expression de r_{jj} , donnée en III.7, montre le processus de régularisation de données de proximité, construit par SIBSON, conduit à définir une distance non euclidienne puisque la distance de KENDALL, définie par $\theta_{jj'}$, est une distance angulaire ([1], [32]).

Si B est le barycentre des points-variables dans le permutoèdre, il a pour k-ième coordonnées dans \mathbb{R}^n

$$b_k = \frac{1}{m} \sum_{j=1}^m y_k^\circ(j) = [\frac{n+1}{2} - \frac{1}{n} \sum_{j=1}^m r_{jk}] / [n(n^2-1)/12]^{1/2} \quad (k = 1, 2, \dots, n)$$

et la norme du vecteur \vec{BG} s'écrit :

$$\|\vec{BG}\|_2^2 = \sum_{k=1}^n b_k^2 = \frac{12}{n(n-1)} \sum_{k=1}^n [\frac{n+1}{2} - \frac{1}{n} \sum_{j=1}^m r_{jk}]^2 = W.$$

expression du coefficient de concordance de KENDALL des classements associés à chacune des variables [1]. Lorsque B et G sont confondus, ceci correspond donc à une situation d'indifférence. A l'opposé, lorsque B appartient à la sphère unité centrée en G, il y a concordance maximale des classements qui sont alors identiques.

La projection des points objets $X^\circ(i)$ sur \vec{BG} traduit ainsi la contribution de l'objet i à la concordance (ou discordance) des jugements. De même la projection sur \vec{BG} du point-variable $Y^\circ(j)$ traduit la déformation du classement moyen par la variable j, si par classement moyen nous entendons le classement construit par la méthode de la somme pondérée des rangs. En effet l'on a [32] :

$$\frac{1}{\sqrt{W}} \langle GX^\circ(i), \vec{GB} \rangle = \frac{1}{\sqrt{W}} b_j$$

Remarque

La généralisation de cette méthode d'analyse à l'étude des jugements de préférence sur p ensembles d'objets $I^1, I^2, \dots, I^q \dots I^p$ est immédiate et permet l'analyse dans $\mathbb{R}^{(I)}$ de telles données, si $I = \bigcup_{q=1}^p I^q$. C'est une des sous méthodes d'analyse des préférences appariées sur $q=1$ des ensembles distincts d'objets (BENZECRI in [18]).

§ 3 - ANALYSE INTERNE : METHODES DE REPRESENTATION FACTORIELLE

Les méthodes de représentation factorielle de données de dominance construisent X et Y en termes de dualité dans un espace euclidien $V = (\mathbb{R}^r, d)$ de dimension r, soit fixée à l'avance, soit déterminée par tâtonnement en itérant la même analyse pour des valeurs consécutives de r. Le point commun de ces méthodes est la formalisation du problème en termes du modèle linéaire ordinal, conduisant à une représentation duale de I et de J.

V-3-1 - Analyse factorielle ordinale

Cette méthode, proposée indépendamment par SHEPARD-KRUSKAL ([58,e], [68,c]) et ROSKAM ([14], ch. 5), suppose que les variables de classement $(j, j \in J)$ sont corrélées et peuvent donc être reconstruites à partir d'un moindre nombre de paramètres qui représentent la structure sous jacente des données sous une forme plus concise et interprétable. C'est donc une analyse factorielle, invariante par transformation isotone des profils δ_j associés à chaque variable et construisant une représentation euclidienne invariante par transformation affine.

Nous disposons d'un tableau de données de préférence Δ défini sur $J \times I$. On se propose de construire un système de points $X = \{X(i), i \in I\}$ dans V et une famille de droites vectorielles $Y = \{Y(j), j \in J\}$ dans V de telle sorte que la projection

$$Z(j,i) = \langle Y(j), X(i) \rangle$$

du point $X(i)$ sur la direction vectorielle $Y(j)$ respecte les contraintes d'isotonie (1).

Par suite le problème est, pour tout j dans J , un problème de régression isotone, puisque si $Y(j)$ est connue, il convient de construire X de telle sorte que $Z(j,.)$ factorise de façon isotone à travers le profil δ_j

$$Z(j,.) = m_j \circ \delta_j$$

Il résulte de la proposition III.8 que la solution ainsi construite est invariante par toute transformation du groupe des similitudes. Par suite on pose les contraintes suivantes :

- i) On impose au système X d'être centré. Par suite, si l'on note X la matrice de $\mathcal{M}_{r,n}(\mathbb{R})$ des coordonnées dans V des vecteurs $X(i)$, ($i \in I$) ; ceci s'écrit :

$$X \mathbf{1}_n = 0$$

où $\mathbf{1}_n$ est le vecteur unité de \mathbb{R}^n .

- ii) La direction vectorielle associée à la variable j est déterminée par le vecteur unitaire $Y(j)$ et l'on pose pour tout j dans J

$${}^t Y(j) Y(j) = 1$$

Assurant ainsi que $Z(j,i)$ définit la projection de $X(i)$ sur la direction vectorielle associée à la variable j .

- iii) Si A est la matrice régulière associée à une transformation affine dans V , en posant :

$$\tilde{X} = AX$$

$$\tilde{Y} = {}^t A^{-1} Y$$

où Y est la matrice de $\mathcal{M}_{r,m}(\mathbb{R})$ de j -ième colonne $Y(j)$, il vient :

$${}^t \tilde{Y} \tilde{X} = {}^t Y A^{-1} A X = {}^t Y X$$

Et donc, bien que \tilde{Y} ne vérifie pas la contrainte ii), la matrice de produits scalaires Z , ($Z \in \mathcal{M}_{m,n}(\mathbb{R})$) vérifie

$$Z = {}^t Y X = {}^t \tilde{Y} \tilde{X}$$

On impose donc aux éléments $X(i)$ d'être réduits.

$$X {}^t X = n I.$$

La construction de X et de Y est alors une simple application des méthodes de regression isotone. En effet, pour tout j dans J lorsque la j -ième ligne $Z_{j.}$ de Z est connue, on construit

$$T_{j.} = E(Z_{j.} / \delta_{j.})$$

qui minimise l'indice de regression isotone

$$\psi_j^2 = \frac{\sum_{i \in I} (Z_{ji} - T_{ji})^2}{\sum_{i \in I} (Z_{ji} - \bar{Z}_j)^2}$$

où

$$\bar{Z}_j = \frac{1}{n} \sum_{i \in I} Z_{ji}$$

Par suite si p_j est une pondération affectée à la variable j , on définit comme indice de regression isotone par morceaux :

$$\psi^2 = \sum_{j \in J} p_j \psi_j^2$$

On construit alors un algorithme qui a chaque itération s franchit trois étapes. Sa description nécessite de poser en notation

$$U = \begin{pmatrix} X(1) \\ X(2) \\ \vdots \\ X(n) \\ Y(1) \\ \vdots \\ Y(m) \end{pmatrix} \quad \text{le vecteur de } \mathbb{R}^q \text{ avec } q = r(n+m)$$

Par abus de langage nous noterons $U = \begin{pmatrix} X \\ Y \end{pmatrix}$.

1ère étape : Minimisation de ψ^2 en Z , sans contraintes.

On utilise une méthode de Gradient et posons

$$U_o^{(s)} = U^{(s-1)} - \alpha^{(s-1)} \left[\frac{d\psi}{dU} \right]^{(s-1)}$$

où $\left[\frac{d\psi}{dU} \right]^{(s-1)}$ est la valeur au point $U^{(s-1)}$ du gradient $\left(\frac{d\psi}{dU} \right) = \begin{pmatrix} \frac{d\psi}{dX} \\ \frac{d\psi}{dY} \end{pmatrix}$

$\frac{d\psi}{dX}$ est le vecteur de \mathbb{R}^m donné, en notations matricielles par :

$$\frac{d\psi}{dX} = \frac{1}{nm\psi} Y(Z-T)$$

de même le vecteur $\frac{d\psi}{dY}$ de \mathbb{R}^{rm} s'écrit en notations matricielles :

$$\frac{d\psi}{dY} = \frac{1}{nm\psi} X^t (Z-T)$$

On construit ainsi le vecteur $U_o^{(s)} = \begin{pmatrix} X^{(s)} \\ 0 \\ Y^{(s)} \\ 0 \end{pmatrix}$ de \mathbb{R}^q .

2ème étape : Respect des contraintes

Si l'on note Q^{-1} la matrice de la transformation affine permettant de centrer en lignes et réduire la matrice X , les contraintes imposées à Y sont alors vérifiées en posant [38,c]

$$Y = Q Y_o D$$

où la matrice diagonale D est définie par

$$D^{-2} = \text{diag} [{}^t Y_o \quad {}^t Q Q Y_o]$$

De plus on a de façon classique pour tout X_o centré :

$$Q = X_o \quad {}^t X_o$$

Par suite on calculera au pas s , $U^{(s)} = \begin{pmatrix} X^{(s)} \\ Y^{(s)} \end{pmatrix}$ avec

$$\left. \begin{aligned} {}^t X^{(s)} &= {}^t X_o^{(s)} [Q^{-1}]^{(s)} \\ Y^{(s)} &= Q^{(s)} Y_o^{(s)} D^{(s)} \end{aligned} \right| \Rightarrow Z^{(s)} = D^{(s)} \quad {}^t Y_o^{(s)} \quad X_o^{(s)}$$

3ème étape : Respect des contraintes d'isotonie (1)

(a) Pour tout j dans J on calcule la regression isotone $T_j^{(s)}$ de $Z_j^{(s)}$ par l'algorithme exposé en II.3.

- (b) **Calcul** de ψ^2 et test d'arrêt
Si $\psi^2 \leq \epsilon = \text{STOP}$
Sinon $s = s+1$, aller à l'étape 1.

Remarque

Cet algorithme itératif nécessite une initialisation, On posera donc $Z^{(0)} = R$ où R est le tableau des rangs centrés réduits étudié au paragraphe précédent. Cette démarche revient, en quelque sorte, à prendre le permutoèdre pour configuration initiale.

Les essais algorithmique d'une telle analyse, montrent que, lorsque les liaisons supposées isotone sont strictement linéaires, la configuration obtenue est très proche de celle obtenue en analyse factorielle linéaire. De plus lorsque ces liaisons sont fortement non linéaires, l'analyse ordinaire ne construit pas de facteurs artificiels, contrairement à l'analyse linéaire. En contrepartie, le coût accru de mise en oeuvre et la sensibilité aux dégénérescences (liées à l'existence de deux sous-groupes I_1 et I_2 de I tels que les classements des objets de I_1 précèdent ceux des objets de I_2) limite l'utilisation de telles méthodes aux problèmes marquant une forte non linéarité.

Lorsque l'on dispose de données de comparaisons par paires, comportant éventuellement des intransitivités, le modèle MDPREF dû à CARROLL [in 15] permet de donner une représentation factorielle de I et J .

A chaque tableau Δ^j , on associe le tableau Σ^j dont le terme générique $\sigma_j(i, \ell)$ est la signature de la relation P_j (Def. III.14). On se propose d'ajuster le modèle

$$Z(j, i) = \langle Y(j), X(i) \rangle$$

défini précédemment, en notant pour tout (i, ℓ) dans $I \times I$

$$U_j(i, \ell) = Z(j, i) - Z(j, \ell)$$

Par suite la relation d'isotonie (1) est vérifiée par le couple (i, ℓ) d'éléments de I , si l'on a

$$\sigma_j(i, \ell) U_j(i, \ell) \geq 0 \quad (\forall j, j \in J)$$

et par suite CARROLL propose de maximiser

$$C = \sum_{j \in J} p_j \frac{[\sum_{i \in I} \sum_{\substack{\ell \neq i \\ \ell \in I}} \sigma_j(i, \ell) U_j(i, \ell)]^2}{\sum_{i \in I} \sum_{\substack{\ell \neq i \\ \ell \in I}} [U_j(i, \ell)]^2}$$

où p_j est une pondération affectée à la variable j , et montre qu'il est équivalent de résoudre le problème aux moindres carrés

$$\|\tilde{\Sigma} - S\|^2 \text{ (Min)}$$

où $\tilde{\Sigma}$ est l'élément de $\mathcal{M}_{m,n}(\mathbb{R})$ de terme générique

$$\tilde{\Sigma}_{ji} = \sqrt{p_j} \sum_{\substack{\ell \neq i \\ (i, \ell) \in I \times I}} [\sigma_j(i, \ell) - \sigma_j(\ell, i)]$$

Le modèle linéaire ordinal conduit donc, en ce cas, à une solution analytique grâce au théorème d'ECKART et YOUNG ([47],[28]).

V.8 - Théorème

Soit Σ un élément de $\mathcal{M}_{m,n}(\mathbb{R})$. La matrice \hat{S} , meilleure approximation de Σ au sens des moindres carrés dans le sous-ensemble \mathcal{S}_r de $\mathcal{M}_{m,n}(\mathbb{R})$ de rang inférieur ou égal à r :

$$\|\Sigma - \hat{S}\|_2 = \text{Min} \{ \|\Sigma - S\|^2 ; S \in \mathcal{S}_r \}$$

est donnée par :

$$\hat{S} = t_U \wedge v$$

où Λ est la matrice diagonale d'ordre r des r valeurs singulières de Σ de plus grand module et U et V sont formes des vecteurs propres à gauche et à droite normalisés, associées à ces valeurs singulières dans la décomposition de Σ .

On peut ainsi déterminer r , puis X et Y en posant, pour un choix donné de r :

$$Y = U, \quad X = \Lambda V.$$

Ce qui assure la condition de normalisation imposée à Y .

Les méthodes ordinales décrites dans ce paragraphe, sont très proches de l'analyse en composantes principales des rangs, préconisée par LEBART. Cependant cette méthode n'est pas à proprement parler ordinale, car le choix des rangs de SPEARMAN comme fonction isotone de Δ , n'a aucun caractère d'optimalité et nous ne sommes nullement assurés de la minimalité de la dimension r de l'espace de représentation V ainsi construit. De plus le choix de la distance de SPEARMAN :

$$d^2(i, i') = \frac{6}{n(n^2-1)} \sum_{j \in J} (r_{ij} - r_{i'j})^2$$

conduit en général à un effet de taille sur le premier axe obtenu, du fait de l'absence de centrage. C'est là un phénomène interprétable à partir des travaux de COOMBS et KAO [44]. Ceux-ci tentèrent de généraliser le modèle de déploiement au cas multidimensionnel en supposant connues les distances interpoints sur chaque axe de classement et non plus seulement les relations d'ordre induites. Une telle méthode conduit à majorer d'un, le nombre de dimensions de la représentation. ROSS et CLIFF [65] montrèrent que ceci est dû au fait qu'il convenait de retenir non pas les distances interpoints, mais leurs carrés. Le premier facteur obtenu s'interprète alors comme la distance à l'origine du barycentre des points idéaux. Il disparaît par double centrage de la matrice conformément aux résultats du paragraphe III.1

§ 4 - ANALYSE EXTERNE DES DONNEES DE DOMINANCE

On suppose ici V et X connus et l'on se propose de construire Y de telle sorte que soient vérifiées les contraintes d'isotonie (1). Nous exposons tout d'abord l'approche de CARROLL [30,c] utilisables pour des données d'échelle intervalle et son extension au traitement de données ordinales. Nous montrons ensuite que les résultats de SHOCKER et SRINIVASAN [72] permettent de construire un analogue purement ordinal à chacun des modèles de CARROLL.

V-4-1 - L'approche métrique de CARROLL

L'étude des données de dominance a retenu l'attention de nombreux chercheurs, dans le cas de données exactes, d'échelle intervalle ([15], [30,c], [56]). Notre exposé est limité à l'étude de données entachées d'erreurs.

Dans ce cas, CARROLL propose une hiérarchie de modèles de complexité décroissante qu'il ajuste au sens des moindres carrés. Alors que le dernier est un modèle vectoriel, les autres utilisent une méthode de regression quadratique pour ajuster la relation

$$(a) \quad \delta_{ji} = \alpha_j d_{ji}^2 + \beta_j + \epsilon_{ji} \quad (i \in I, j \in J)$$

où α_j et β_j sont des constantes à estimer et ϵ_{ji} un terme d'erreur. Les modèles sont distingués suivant l'expression de la matrice W_j ($W_j \in \mathcal{M}_{r,r}(\mathbb{R})$) définissant :

$$(b) \quad d_{ji}^2 = {}^t(Y(j)-X(i)) W_j (Y(j)-X(i))$$

L'expression (a) s'écrit alors :

$$(c) \quad \delta_{ji} = {}^tX(i) A^j X(i) + B^j X(i) + c_j + \epsilon_{ji}$$

avec :

$$(d-1) \quad A^j = \alpha_j W_j$$

$$(d-2) \quad B^j = -2\alpha_j {}^t Y(j) W_j$$

$$(d-3) \quad c_j = \beta_j + \alpha_j {}^t Y(j) W_j Y(j)$$

où $Y(j)$ est l'inconnue du problème et X est connue.

i) Le modèle général de déploiement :

On associe à chacune des variables j dans J une matrice W_j qui autorise une pondération négative de certains des axes obtenus par rotation des axes originels de V . On autorise ainsi l'existence de points anti-idéaux, traduisant un "pôle de répulsion" plutôt qu'un pôle de préférence. L'équation (c) s'écrit :

$$\delta_{ji} = \sum_{k=1}^r \sum_{\ell=1}^r A_{k\ell}^j x_k(i) x_\ell(i) + \sum_{k=1}^r B_k^j x_k(i) + c_j + \epsilon_{ji}$$

et le problème est résolu par un algorithme de regression multiple avec pour variables indépendantes les $x_k(i)$ et pour variables pseudo-indépendantes les $x_k(i) x_\ell(i)$. On obtient ainsi une estimation $(\hat{A}^j, \hat{B}^j, \hat{c}_j)$ de (A^j, B^j, c_j) . On en déduit l'estimation cherchée grâce à (d-2) :

$${}^t \hat{Y}(j) = - \frac{1}{2} \hat{B}^j [\hat{A}^j]^{-1}$$

et la matrice \hat{A}^j permet de construire la représentation ${}_j X$ déduite de X , associée à la variable j par décomposition spectrale de \hat{A}^j .

ii) Le modèle de déploiement sans rotation des axes

On associe à chacune des variables j dans J une matrice diagonale W_j . On suppose l'orientation des axes commune à toutes les variables j de J , et l'équation (c) s'écrit de façon simplifiée

$$\delta_{ji} = \sum_{k=1}^r A_{kk}^j x_k^2(i) + \sum_{k=1}^r B_k^j x_k(i) + c_j + \varepsilon_{ji}$$

ce modèle nécessite la détermination de $(2r+1)$ inconnue au lieu de $\frac{r(r-1)}{2} + 1$ et les variables pseudo indépendantes sont les $x_k^2(i)$. Le modèle, moins sophistiqué que le précédent, est souvent mieux adapté aux problèmes concrets et plus informatif car d'une interprétation plus aisée.

iii) Le modèle de déploiement simple

C'est un cas particulier du précédent, obtenu en imposant aux termes diagonaux w_k de prendre l'une des valeurs $+1$ ou -1 et ceci de façon identique pour tous les éléments j dans J . Par suite l'équation (c) s'écrit :

$$\delta_{ji} = \alpha_j \sum_{k=1}^r w_k x_k^2(i) + \sum_{k=1}^r B_k^j x_k(i) + c_j + \varepsilon_{ji}$$

et le problème de regression est simplifié puisque

$$t(i) = \sum_{k=1}^r w_k x_k^2(i)$$

est considéré comme une variable pseudo-indépendante unique dans l'équation de regression et l'on obtient de façon directe l'estimation de r paramètres α_j et B_k^j . Les coordonnées de $\hat{Y}(j)$ sont données dans (d-2) par

$$\hat{y}_k(j) = - \frac{1}{2} \frac{\hat{B}_k^j}{\alpha_j w_k}$$

iv) Le modèle vectoriel

Dans ce modèle, $Y(j)$ représente, pour tout j dans J , une direction vectorielle dans V , issue du barycentre des $X(i)$. L'ordre des projections des $X(i)$ sur $Y(j)$ doit alors vérifier les contraintes d'isotonie (1). L'équation de regression s'écrit en ce cas :

$$\delta_{ji} = \alpha_j t_{Y(j)} X(i) + \beta_j + \varepsilon_{ji}$$

où $Y(j)$ est le vecteur unitaire de la direction vectorielle associée à la variable j .

Cette équation s'écrit :

$$\delta_{ji} = t_{W^j} X(i) + \beta_j$$

où les paramètres β_j et

$$W^j = \alpha_j Y(j)$$

sont estimés en résolvant cette équation de régression linéaire. On en déduit $\hat{Y}(j)$ lorsque l'on connaît le vecteur \hat{W}^j

$$\hat{y}_k(j) = \frac{\hat{w}_k^j}{\|\hat{W}^j\|}$$

ou $\| - \|$ est la norme euclidienne usuelle dans \mathbb{R}^T .

V-4-2 - L'approche semi métrique de CARROLL

Les différents modèles étudiés par CARROLL sont résolus par régression quadratique. Lorsque les données sont d'échelle ordinale, les résultats du chapitre II conduisent à poser le modèle de régression isotone :

$$(**) \quad m_j(\delta_{ji}) = d_{ji}^2 + \epsilon_{ji}$$

Pour tous j dans J . L'expression (b) permet de construire ainsi chacun des quatre modèles précédents, en itérant un algorithme en deux phases, jusqu'à convergence :

1ère phase

Ajustement du modèle de régression métrique

$$(*) \quad z_{ji}^{(s-1)} = f_j(d_{ji}^{(s)}) \quad (i \in I, j \in J)$$

où f_j est l'expression de la fonction de regression associée à chacun des modèles ci-dessus avec la contrainte supplémentaire : $\alpha_j > 0$.

2ème phase

Construction, pour tout j dans J , de la regression isotone $Z_j^{(s)}$ de l'estimation $\hat{Z}_j^{(s-1)}$ de $Z_j^{(s-1)}$ construite en phase 1.

L'algorithme est interrompu lorsque l'on ne note une amélioration ni des coefficients de regression du modèle (\star), ni des valeurs de la regression isotone $Z_j^{(s)}$.

Le choix du modèle adapté aux données traitées est réalisable par des méthodes inférencielles classiques, lorsque les données sont d'échelle inter-
valle. Lorsqu'elles sont d'échelle ordinale, aucune méthode ne permet de choisir le modèle le plus informationnel, le test de FISHER n'étant plus valide.

V-4-3 - Approche purement ordinale

L'utilisation de techniques de regression polynomiale a permis à CARROLL de construire ses quatre modèles. Pourtant la relation ($\star\star$) montre qu'il est possible d'aborder ce problème en termes de regression isotone.

La mise en oeuvre ordinale du modèle vectoriel est une conséquence immédiate du modèle linéaire ordinal étudié en II.4. Il suffit en effet d'appliquer ce modèle pour chacun des éléments j dans J . L'utilisation de l'approche de SRINIVASAN et SHOCKER, étudiée en IV.3 conduit, en pratique, à un algorithme plus efficace et moins coûteux que l'approche semi-métrique de CARROLL.

Le modèle général de déploiement de CARROLL nous semble inadapté aux données ordinales puisque alors X est invariant par rotation. Aussi donnons nous un modèle, permettant de construire une solution ordinale des deux autres modèles de déploiement proposés par CARROLL. Notre étude du modèle de SRINIVASAN et SHOCKER [72, a] exclue, dans un premier temps, la possibilité d'exaequos. Ce cas est étudié dans un deuxième temps.

Nous supposons que pour tout j dans J , la distance d_{ji} est donnée par

$$(b) \quad d_{ji}^2 = {}^t(Y(j)-X(i)) W^j (Y(j)-X(i))$$

ou la matrice carrée d'ordre r : W^j est supposée diagonale. Par suite si l'on impose la contrainte

$$(e) \quad W_k^j \geq 0 \quad (\forall k ; k \in \{1,2,\dots,r\}, \forall j ; j \in J)$$

on est assuré que la distance d_{ji} est fonction décroissante du classement de l'objet i pour la variable j .

Pour tout j dans J le modèle est déterminé lorsque l'on connaît W^j et $Y(j)$. La méthode de construction utilisée est, de même qu'en IV.3, identique pour tout j dans J . Aussi l'indice j sera t'il négligé lorsque ceci ne conduit pas à une ambiguïté.

Notations

Soit M le sous-ensemble de $A_j \times A_j$, défini pour la variable j étudiée par $M = \{(a,a') ; (a,a') \in A_j \times A_j ; \sigma_p(a,a') \geq 0\}$ que l'on identifie à $\{1,2,\dots,m\}$. Nous noterons de plus k^j l'injection canonique de M dans $A_j \times A_j$ et adoptons les notations de IV.3, ainsi notera t'on

$$U_t = d_a^2 - d_{a'}^2$$

Pour tout t dans M vérifiant

$$k(t) = (a,a').$$

SRINIVASAN et SHOCKER se proposent de construire la solution (W^*, Y^*) par minimisation de la fonction d'écart définie en IV-3 par :

$$L(1) = \frac{\sum\{|U_t| ; t \in M_-\}}{\sum\{|U_t| ; t \in M\}} = [1 + \frac{F^+(1)}{F^-(1)}]^{-1}$$

où $F^-(1) = \sum\{|U_t| ; t \in M_-\}$ si $M_- = \{t ; t \in M \quad U_t > 0\}$

$F^+(1) = \sum\{|U_t| ; t \in M_+\}$ si $M_+ = \{t ; t \in M \quad U_t < 0\}$

l'équivalence de notre formulation à celle des auteurs est assurée par la

V-9 - Proposition [72,a]

Lorsqu'elle existe la solution (W^*, Y^*) du problème

$$P_1 \quad \left\{ \begin{array}{l} \frac{F^-(1)}{F^+(1)} = (\text{Min}) \\ F^+(1) - F^-(1) = h \quad ; \quad k \in \mathbb{R} \quad \{0\} \\ W \geq 0 \end{array} \right.$$

est invariante par homothétie sur W , non triviale et indépendante de la valeur de h .

On imposera donc, dans la suite, l'égalité à 1 de h . Par suite, si l'on note

$$\zeta_k(i, \ell) = x_k^2(i) - x_k^2(\ell)$$

$$\xi_k(i, \ell) = x_k(i) - x_k(\ell)$$

et V le vecteur de \mathbb{R}^F de k -ième composante

$$v_k = y_k w_k$$

il vient :

$$\begin{aligned}
 1 &= F^+(1) - F^-(1) = \sum_{t \in M} U_t \\
 &= \sum \left\{ \sum_{k=1}^r w_k (x_k(i) - y_k)^2 - \sum_{k=1}^r w_k (x_k(l) - y_k)^2 ; ((i,j), (l,j)) \in M \right\} \\
 &= \sum \left\{ \sum_{k=1}^r w_k \zeta_k(i,l) + \sum_{k=1}^r v_k \xi_k(i,l) ; ((i,j), (l,j)) \in M \right\} \quad (***) \\
 &= {}^t_A W. + {}^t_D V
 \end{aligned}$$

ou A (resp. D) est le vecteur de \mathbb{R}^r de k-ième coordonnée :

$$A_k = \sum \{ \zeta_k(i,l) ; ((i,j), (l,j)) \in M \}$$

$$D_k = \sum \{ \xi_k(i,l) ; ((i,j), (l,j)) \in M \}$$

de plus la relation (***) donne l'expression de $F^-(1)$:

$$(***) F^-(1) = \sum \{ \text{Max}(0, \sum_{k=1}^r w_k \zeta_k(i,l) + \sum_{k=1}^r v_k \xi_k(i,l)) ; ((i,j), (l,j)) \in M \}$$

Remarque

L'expression (***) montre le caractère technique de cette contrainte. En effet, puisque Y n'est pas contraint à être positif, il en est de même pour V. Le problème admet donc toujours pour solution admissible, le couple (W,V) tel que :

$$V_k = W_k = 0 \quad \forall k, k \neq k_0, \quad k = 1, 2, \dots, r$$

pour k_0 tel que $D_{k_0} \neq 0$

$$W_{k_0} = 1$$

$$V_{k_0} = (1 - A_{k_0}) / D_{k_0}$$

On peut alors énoncer la :

V-10 - Proposition

Une condition nécessaire et suffisante d'existence d'une solution (W,Y) du problème P₁ est qu'il existe une solution au problème

$$P_2 \left\{ \begin{array}{l} W \geq 0 \\ r_{AW} + t_{DV} = 1 \\ F^-(1) \text{ donnée par } (***) \text{ est minimale.} \end{array} \right.$$

Démonstration

Seule la condition suffisante pose un problème. Nous devons distinguer trois cas :

i) $w_k > 0$
On pose alors $y_k = \frac{v_k}{w_k}$

ii) $w_k = v_k = 0$.
La dimension k n'intervient pas dans le classement des objets de I pour la variable étudiée et l'on annule donc y_k .

iii) $w_k = 0, v_k \neq 0$
La valeur de y_k est alors infinie. Ce cas, très intéressant, correspond au modèle vectoriel. En effet, si y_k tend vers l'infini de telle sorte que $y_k w_k$ tende vers une valeur finie v_k , il résulte que

$$d_{ji}^2 - d_{jl}^2 = w_k (x_k^2(i) - x_k^2(l)) - 2w_k y_k (x_k(i) - x_k(l))$$

tend vers $-2 v_k (x_k(i) - x_k(l))$.

On construit le modèle mixte, déterminé par le vecteur T : tel que $\sum_{k=1}^r t_x x_k(i)$ traduise le classement de l'objet i pour la variable étudiée avec ici

$$t_k = + \frac{1}{2} v_k$$

Ainsi a-t-on i) un modèle de distance sur un sous-espace V_1 de V tel que

$$w_k > 0 \text{ ou } v_k = w_k = 0$$

ii) un modèle vectoriel sur le complémentaire de V_1 dans V sur lequel

$$y_k = \text{signe}(v_k) \times (+\infty).$$

Le problème P_2 est résolu de façon classique en introduisant les variables positives d'écart $Z(i,\ell)$ et résolvant le programme linéaire

$$P_3 \left\{ \begin{array}{l} W \geq 0 \quad Z \geq 0 \quad Z \in \mathbb{R}^p \quad p = \frac{n(n-1)}{2} \\ t_{AW} + t_{DV} \geq 0 \\ \sum_{k=1}^r [\zeta_k(i,\ell)w_k + \xi_k(i,\ell)v_k] + Z(i,\ell) \geq 0 \quad (V(i,\ell) ; ((i,j),(\ell,j)) \in M) \\ \sum \{Z(i,\ell) ; ((i,j),(\ell,j)) \in M\} = F(\text{Min}) \end{array} \right.$$

qui admet une solution optimale finie puisque F est bornée inférieurement par zéro et puisque le simplexe défini par les contraintes de P_3 est non vide. En effet, le triplet $(W^\circ, V^\circ, Z^\circ)$ défini par

$$V_k^\circ = W_k^\circ = 0 \quad (\forall k ; k \neq k_0, k \in \{1,2,\dots,r\})$$

où k_0 est un élément de $\{1,2,\dots,r\}$ pour lequel D_{k_0} est non nul,

$$V_{k_0}^\circ = (1 - A_{k_0}) / D_{k_0}$$

$$W_{k_0}^\circ = 1$$

$$Z^\circ(i,\ell) = \text{Max}\{0, -[\frac{1 - A_{k_0}}{D_{k_0}} \xi_{k_0}(i,\ell) - \zeta_{k_0}(i,\ell)]\} \quad ((i,j),(\ell,j)) \in M$$

est une solution réalisable de P_3 , existant toujours si X n'est pas une solution dégénérée, auquel cas le problème de construction de Y ne se pose pas.

Remarques

La résolution du problème P_3 permet de déterminer les éléments (a, a') dans M qui violent les contraintes d'isotonie (1) car alors la variable artificielle correspondante est non nulle, et sa valeur donne sa contribution à F^* . Ce fait est intéressant en pratique, pour déceler les intransitivités contenues dans les données.

L'unicité de la solution (W^*, V^*) n'est pas assurée. S'il en existe une infinité réalisant le minimum F avec $Z^* \equiv 0$, on prendra pour solution celle qui maximise la distance minimum aux hyperplans frontières de la facette décrite par (W^*, V^*) . Dans le cas contraire, $Z^* \neq 0$ et l'on choisira la solution qui minimise le nombre de violations des contraintes d'isotonie.

Lorsque les exaequos sont admis, on note M^1 (resp. M^2) le sous-ensemble de M composé des jugements tranchés (resp. des exaequos) où :

$$M^1 = \{(a, a') \in M ; d_a^2 - d_{a'}^2 \geq \epsilon \}$$

$$M^2 = \{(a, a') \in M ; d_a^2 - d_{a'}^2 < \epsilon \}$$

où ϵ est un nombre réel positif arbitrairement fixé. On définit alors :

$$F_1^-(1) = \Sigma \{|U_t - \epsilon| ; t \in M_- \cap M_1\}$$

$$F_1^+(1) = \Sigma \{|U_t - \epsilon| ; t \in M_+ \cap M_1\}$$

et l'on associe à M^2 l'ensemble

$$M^3 = \{(a, a') ; (a, a') \in M^2 \text{ ou } (a', a) \in M^2\}$$

qui permet de n'imposer aucune contrainte aux éléments de M^3

$$F_3^-(1) = \Sigma \{ |U_t| ; t \in M^3 \}$$

$$F_3^+(1) = \Sigma \{ |U_t| ; t \in M^3 \}$$

et l'on minimise comme précédemment l'indice d'écart $L(1)$ sur la réunion de M^1 et M^3 . le problème P_3 s'écrit alors :

$$(P'_3) \left\{ \begin{array}{l} W \geq 0 \quad Z \geq 0 \\ \\ t_{AW} + t_{DV} = 1 + \text{card}(M^1)\epsilon \\ \\ \sum_{k=1}^r \{ \zeta_k(i,l)w_k + \xi_k(i,l)v_k \} + Z(i,l) \geq \epsilon \quad (\forall(i,l) ; ((i,j), (l,j)) \in M^1) \\ \\ \sum_{k=1}^r \{ \zeta_k(i,l)w_k + \xi_k(i,l)v_k \} + Z(i,l) \geq 0 \quad (\forall(i,l) ; ((i,j), (l,j)) \in M^3) \\ \\ \Sigma \{ Z(i,l) ; ((i,j), (l,j)) \in M^1 \cup M^3 \} = F(\text{Min}) \end{array} \right.$$

Remarques

- i) L'existence d'une solution de (P'_3) est assurée de façon identique au cas précédent.
- ii) L'introduction d'un modèle ϵ de "confusion" est identique à l'approche de de LEEUW (cf. III.4). Sa détermination arbitraire pose cependant des problèmes en pratique.

Cet algorithme exposé dans le cadre de validité du modèle II de CARROLL, est rendu plus performant en résolvant le problème linéaire dual [72,a].

B I B L I O G R A P H I E

§ 1 - L I V R E S

- [1] BARBUT, M. Techniques ordinales en analyse des données.
FREY, L. Algèbre et combinatoire - Hachette Université 1971
- [2] BARLOW, R.E. Statistical inference under order restrictions
BARTOLOMEW, D.J. Wiley - 1972.
BREMNER, J.M.
BRUNK, H.D.
- [3] BENZECRI, J.P. L'Analyse des données - 2 tomes - Dunod 1973
- [4] BLUMENTHAL, L.M. Theory and applications of distance geometry -
CHELSEA Publi. 1970
- [5] COOMBS, C. H. A theory of data - Wiley 1964.
- [6] DEGENNE, A. Techniques ordinales en analyse des données :
statistiques - Hachette Université 1972.
- [7] DUBOS, J. Liaisons stochastiques en économie. Dunod 1971
- [8] FREEMAN, L.C. Elementary applied statistics - Wiley 1965
- [9] GREEN, P.E. Multidimensionnal scaling and related techniques
CARMONE, F.J. in marketing analysis. Allyn and Bacon 1972
- [10] GREEN, P.E. Applied multidimensional scaling.
RAO, U.R. Holt, Rinehart and Winston Inc. 1972
- [11] GREEN, P.E. Multiattribute decisions in marketing
WIND, Y. Dryden press. Hinsdale Illinois 1973
- [12] KENDALL, M.G. Rank correlation methods. Griffin Londres 1962
- [13] LERMAN, I.C. Les bases de la classification automatique
Gauthier Villars 1970.
- [14] ROSKAM, E.E.C. Metric analysis of ordinal data in psychology
Van Voorschotten 1968.
- [15] SHEPARD, R.N. Multidimensionnal scaling : theory and appli-
ROMNEY, A.K. cations in the behavioral sciences - 2 tomes
NERLOVE, S.B. Seminar Press 1972

- [16] THURSTONE, L.L. Multiple factor analysis - University of Chicago Press 1947.
- [17] TORGERSON, W.S. Theory and methods of scaling - Wiley 1958.

§ 2 - S E M I N A I R E S et T H E S E S

- [18] Aix-en-Provence Ordre totaux finis - Mouton Gauthier Villars 1971
- [19] BESSON, M. Quelques méthodes d'analyse multicritère - thèse de 3ème cycle Stat. Maths. Paris mars 1972
- [20] DAVY de VIRIVILLE, M L'analyse des proximités. Thèse 3ème cycle ISUP Paris VI juin 1972.
- [21] FENELON, J.P. Deux contributions à une programmation d'analyse des données ISUP Paris VI février 1973
- [22] LAURENT, J.P. Un algorithme dual pour le calcul de la distance entre deux convexes. Séminaire d'analyse numérique de Grenoble 27-02-1975 (N° 220)
- [23] LINGOES, J.C. Recent computational advances in nonmetric methodology for behavioral sciences. International computation symposium Rome 1966.
- [24] MAMAIA 1970 Conférence "Mathematics in the archeological and historical sciences" édité par F.R. HODSON, D.G. KENDALL et P. TAUTU - Edimbourg University press 1971.
- [25] ROME 1966 Théorie des graphes ; Journées internationales Dunod 1967.

§ 3 - P U B L I C A T I O N S M U L T I G R A P H I E E S

- [26] BESSON, M. Une généralisation de la notion de rang pour les ensembles finis préordonnés. Ecocentre Mac Gregor Paris sept. 1973.
- [27] BOUROCHE, J.M. Les modèles indscal et idioscal - Méthodes et utilisation. Metra - direct. scient. Note de travail 192 octobre 1973.
- [28] BOUROCHE, J.M. MEDVEDEV, P.

- [29] BOUROCHE, J.M
LAYE, P. Analyse des préférences - Modèles PREFMAP I et II. Metra direct-scient; Note de travail 190 octobre 1973.
- [30] CARROLL, J.D. BELL laboratories. Murray Hill - New Jersey 07974
a] Models and Methods for three way multidimensional scaling.
b] Introduction to INDSCAL .
c] Models and methods for multidimensional preference scaling .
d] Individual differences and multidimensional scaling.
- [31] DUSSAIX, A.M Quelques méthodes d'analyse et de structuration de données cubiques. Metra direct. scient. Note de travail 191 - octobre 1973.
- [32] JACQUET LAGREZE, E Méthodes et programmes en analyse des données ordinales. Metra direct.scient. Synthèse et formation 81 - Décembre 1973.
- [33] DE LEEUW, J Department of data theory University of Leiden. The Netherland .
a] Non metric multidimensional scaling RN01068
b] The linear non metric model RN00369
c] The positive orthant method for non metric multidimensional scaling RN 00170
d] The euclidean distance model RN00270
e] Bounds for STRESSFORM 1 (part.1 RBO0173 - part.2 RBO0473)
f] Bounds for normalised phi RB00373
- [34] LINGOES, J.C
ROSKAM, E.E.C. A mathematical and empirical study of two multidimensional scaling algorithms.
- [35] Mc GEE, V.E. Speech perception and the developpement of appropriate non metric analytical techniques. Final report on office of naval research. Department of the Navy Washington 1969.

§ 4 - A R T I C L E S

- [36] BEALS, R Foundations of multidimensional scaling
KRANTZ, D.H. Psychological review - 1968 - vol 75, N°2 pp.
TVERSKY, A. 127 - 142.
- [37] BENNETT, R.S. The intrinsic dimensionality of signal
collections. IEEE transactions on information
theory vol. I.T.15 , N°5 sept. 1969
- [38] BENNET, J.F. Multidimensional unfolding. Determining the
HAYS, W.S. dimensionality of preference data. Psychome-
trika 1960, Vol.25 pp.27-43.
- [39] BENZEKRI J.P. Sur l'analyse factorielle des proximités
Publication de l'ISUP I 1964 Vol.13 et II. 1965
Vol.14.
- [40] BERNARD, G. Douze méthodes d'analyse multicritère
BESSON, M.L. R.I.R.O. V, 3 - 1971 pp.19-66.
- [41] BOUROCHE, J.M. Analyse des proximités et analyse des préfé-
rences. Metra vol X, n°4, 1971.
- [42] DE CANI, J.S. Maximum likelihood paired comparison ranking
a) by linear programming - Biometrika 1969 -56-2 pp
b) A branch and bound algorithm for maximum like-
lihood paired comparison ranking - Biometrika
1972 - 59-1 pp.131-135.
- [43] CARROLL, J.D. Analysis of individual differences in multi-
dimensional scaling via a N.way generalisation
of Eckart-Young decomposition.
Psychometrika 1960 - 25-3 pp. 283-319.
- [44] COOMBS, C.H. On the connection between factor analysis and
KAO, R.C. multidimensional unfolding. Psychometrika 1972
37-3 pp. 219-231.
- [45] COOPER, L.C. A new solution to the additive constant problem
in the metric multidimensional scaling.
Psychometrika 1972 - 37-3 pp.311 - 322.
- [46] DEGENNE, A. Introduction à l'analyse de similitude.
VERGES, P. Revue française de sociologie 1973 XIV pp.471-
512.

- [57] KLAHR, D. A monte carlo investigation of the statistical significance of Kruskal's non metric scaling procedure.
Psychometrika 1969 vol. 34 N°3. pp.319-330.
- [58] KRUSKAL, J.B.
a) Multidimensional scaling byoptimising goodness of fit to a non-metric hypothesis.
Psychometrika 1964, vol.29, I. pp.1-27, II pp. 115-129.
b) Analysis of factorial experiments by estimating monotone transformations of the data.
Journal of the royal statistical society B, 1965, 27 pp. 251-263.
c) Monotone regression : continuity and differentiability properties.
Psychometrika 1971 - 36 pp. 57-62.
d) avec Geometrical models and badness of fit functions.
CARROLL, J.D. krishnaiah P.R (ed.) Acc. press 1969 pp.639-670
e) avec A non metric variety of linear factor analysis
SHEPARD, R.N. Psychometrika, 1974 39-2 pp. 123-158.
- [59] LINGOES, J.C. Some boundary conditions for the monotone analysis of symetric matrices. Psychometrika 1971 36-2 pp. 195-203.
- [60] LORD, R.D. The distribution of distances in a hypersphere.
Annals of mathematical statistics.1954-25 pp. 794-798.
- [61] LUCE, R.D. Simultaneous conjoint measurement.
TUKEY, J.W. Journal of mathematical psychology 1964,1 pp. 1-27
- [62] Mc DONALD, R.P. Nonlinear factor analysis.
Psychometrika monographs N°15 - publ. psychometric society.
- [63] Mc GEE, V.E. The multidimensional analysis of elastic distances.
a] British Journal of math et stat. psychology 1966-19-2 pp. 181-196.

- [63] b) Multidimensional scaling of N sets of similarity measures. Multivariate behavioral research 1968-3 pp. 233-248.
- [64] MESSICK, S.J. The additive constant problem in multidimensional scaling. Psychometrika 1956 vol. 21 N°1 pp. 1-17.
- [65] ROSS, J.
CLIFF, N. A generalisation of the interpoint distance model. Psychometrika 1960 Vol. 25 N°3 pp.219-231.
- [66] SCHÖNEMANN. P. H. On metric multidimensional unfolding. Psychometrika 1970-35-3 pp. 349-366.
- [67] SCHWARTZ, H.R. The method of overrelaxation for $(A-\lambda B)X = 0$ Numer. Math. 1974-23 pp. 135-151.
- [68] SHEPARD, R.N.
a) Multidimensional scaling with unknown distance function. Psychometrika 1962,27,I pp. 125-140
II - pp. 219-246.
b) Metric structures of ordinal data. Journal of mathematical psychology 1966-3-2 pp. 287-315.
c) Representation of structure in similarity data. Psychometrika 1974-39-4 pp. 373-422.
d) avec KRUSKAL, J.B. Nonmetric methods for scaling and for factor analysis. American psychologist 1964-19 pp.557-558.
e) avec CARROLL, J.D. Paramtric representation of non linear data structures. in Krishnaiah 1966. Accademic Press pp. 561-592.
f) avec CUNNINGHAM, J.P. Monotone mapping of similarities into a general metric space. Journal of mathematical psychology 1974-11-4 pp. 335-363.
- [69] SHERMAN, C.R. Non metric multidimensional scaling. A monte carlo study of the basic parameters. Psychometrika 37-3 1972 pp. 323-355.
- [70] SIBSON, R. Order invariant methods for data analysis. Journal of Royal statis. society A 1972.5 pp. 311-349.

- [71] SPENCE, I. A monte Carlo evaluation of three multidimensional scaling algorithms. *Psychometrika* 1972 37-4 pp. 461-486.
- [72] SRINIVASAN, V.
SHOCKER, A.D.
a)
b)
- Linear programming techniques for multidimensional analysis of preferences. *Psychometrika* 1973. 38-3 pp. 337-369.
- Estimating the weights for multiple attributes in a composite criterion using pairwise judgments. *Psychometrika* 1973 -38-4 pp. 473-493.
- [73] STENSON, H.H.
KNOLL, R.L.
- Goodness of fit for random rankings in Kruskal's nonmetric scaling procedure. *Psychological bulletin* 1969 -71-2 pp. 122-126.
- [74] TORGERSON, W.S.
- Multidimensional scaling of similarity. *Psychometrika* 1965-30-4 pp. 379-393.
- [75] TRUNK, G.V.
- Statistical estimation of the intrinsic dimensionality of data collections. *Information and control* 1968-12 pp. 508-525.
- [75, a] TUCKER, L.R.
- Relations between multidimensional scaling and three mode Factor analysis. *Psychometrika* 1972 37-1 pp. 3-27.
- [75, b] TUCKER, L.R.
MESSICK, S.
- "An individual difference model for multi dimensional scaling". *Psychometrika* 1963 - Vol.28 pp. 333-367.
- [76] TVERSKY, A
a)
b) avec
KRANTZ, D.H.
- Intransitivity of preferences. *Psychological review* 1969 - 76-1 pp. 31-48.
- The dimensional representation and the metric structure of similarity data. *Journal of mathematical psychology* 1970-7- pp. 572-596.
- [77] VAN EEDEN, C
- Maximum likelihood estimation of partially and completely ordered parameters. *Proc. Akademisch van Wetenschappen. A*; 1957. I. pp. 128-136 et II pp. 506-612.
- [78] WOLD, H.
- Estimation of principal components and related

De même, les concepts de statistique mathématique utilisés dans le texte sont définis et étudiés dans :

- [84] BARRA, J.R. Notions fondamentales de statistique mathématique DUNOD -Paris 1971.
- [8] ULMO, J Eléments de décision statistique
 BERNIER, J. PUF Paris 1973.

models by iterative least squares. In
Krishnaiah p.r. (ed.) Acc. Press 1966 pp.391-
420.

[79] YATES, A. Non metric individual differences multidimen-
sional scaling with balanced least squares
monotone regression. Paper presented to the
psychometric society meeting. March 1972.

[80] YOUNG. F.W. A fortran IV program for nonmetric multidi-
a) mensional scaling TORSCA 9. Behavioral science
1968-13- pp. 343-344.

b) Non metric multidimensional scaling. Recovery
of metric information. Psychometrika 1970
35-4 pp. 455-473.

Remarque :

Aucune référence n'est donnée quant aux méthodes numériques usuelles
citées dans le texte. Une introduction synthétique aux méthodes utilisées
en statistique, ainsi qu'une bibliographie détaillée sont données par :

[81] CHAMBERS, J.M. Fitting non linear models : Numerical techni-
ques - Biometrika 1973-60 pp. 1-13.

Le lecteur intéressé peut se référer aux manuels :

[82] GASTINEL, N. Analyse numérique linéaire. Herman Paris 1966.

[83] KUNZI, H.P. La programmation non linéaire. Gauthier Villars
KRELLE, W. Paris 1969.

Dernière page d'une thèse

VU

Grenoble, le 2/6/75

Le Président de la thèse



J.R. BARRA

Vu, et permis d'imprimer,

Grenoble, le

Le Président de l'Université
Scientifique et Médicale

